Systematic Entomology (2002) **27**, 361–382

Phylogeny of Anophelinae (Diptera: Culicidae) based on nuclear ribosomal and mitochondrial DNA sequences

M. A. M. SALLUM*, T. R. SCHULTZ†, P. G. FOSTER‡, K. ARONSTEIN§, R. A. WIRTZ¶ and R. C. WILKERSON§

*Nucleo de Pesquisas Taxonômicas e Sistemáticas em Entomologia Médica, Departamento de Epidemiologia, Faculdade de Saúde Pública, Universidade de São Paulo, Brazil, † Department of Systematic Biology, National Museum of Natural History, Smithsonian Institution, Washington, DC, U.S.A., ‡Department of Zoology, The Natural History Museum, London, U.K., §Department of Entomology, Walter Reed Army Institute of Research, Smithsonian Institution, Washington, DC, U.S.A., ¶Entomology Branch, Division of Parasitic Diseases, National Center for Infectious Diseases, Atlanta, GA, U.S.A.

Abstract. Phylogenetic relationships among thirty-two species of mosquitoes in subfamily Anophelinae are inferred from portions of the mitochondrial genes COI and COII, the nuclear 18S small subunit rRNA gene and the expansion D2 region of the nuclear large subunit 28S rRNA gene. Sequences were obtained from the genera *Anopheles*, *Bironella* and *Chagasia*. Representatives of all six subgenera of *Anopheles* were included: *Anopheles*, *Cellia*, *Kerteszia*, *Lophopodomyia*, *Nyssorhynchus* and *Stethomyia*. Using parsimony and maximum likelihood methods, various combinations of these DNA sequence data were analysed separately: 18S, 28S, combined 18S and 28S, combined COI and COII, and combined 18S, 28S, COI and COII ('total evidence'). The combined rDNA data contain strong phylogenetic signal, moderately to strongly supporting most clades in MP and ML analyses; however, the mtDNA data (analysed as either nucleotide or amino acid sequences) contain little phylogenetic signal, except for relationships of very recently derived groups of species and, at the deepest level, for the monophyly of Anophelinae. The paraphyly of *Anopheles* relative to *Bironella* is confirmed by most analyses and statistical tests. Support for the monophyly of subgenera *Anopheles*, *Cellia*, *Kerteszia* and *Nyssorhynchus* is indicated by most analyses. Subgenus *Lophopodomyia* is reconstructed as the sister to *Bironella*, nested within a clade also containing *Nyssorhynchus* and *Kerteszia*. The most basal relationships within genus *Anopheles* are not well resolved by any of the data partitions, although the results of statistical analyses of the rDNA data (S-H-tests, likelihood ratio tests for monophyly and Bayesian MCMC analyses) suggest that the clade consisting of *Bironella*, *Lophopodomyia*, *Nyssorhynchus* and *Kerteszia* is the sister to the clade containing *Cellia* and *Anopheles*.

Introduction

Subfamily Anophelinae originally included three genera: *Anopheles* Meigen, which is almost world-wide in distribution; *Bironella* Theobald, found in the Australasian Region, mainly New Guinea; and the Neotropical *Chagasia* Cruz (Knight & Stone, 1977). Currently, Anophelinae is divided into two genera, *Anopheles* and *Chagasia* (Sallum *et al.*, 2000), and is widely agreed to be the basal clade within Culicidae (Pawlowski *et al.*, 1996; Besansky & Fahey, 1997; Miller *et al.*, 1997; Harbach & Kitching, 1998). The phylogenetic relationships among members of the subfamily have been investigated on the basis of morphological and molecular characters (Besansky & Fahey, 1997; Foley *et al.*, 1998; Harbach & Kitching, 1998; Sallum *et al.*, 2000; Krzywinski *et al.*, 2001a,b). The results of the various studies disagree, in particular with the hypothesis of paraphyly of genus *Anopheles* relative to genus *Bironella*

Correspondence: Maria Anice Mureb Sallum, Departamento de Epidemiologia, Faculdade de Saúde Pública, Universidade de São Paulo, Av. Dr Arnaldo 715, CEP 01246-904, São Paulo, SP, Brazil. E-mail: masallum@usp.br

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
2002		**00-00-2002 to 00-00-2002**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Phylogeny of Anophelinae (Diptera: Culicidae) based on nuclear ribosomal and mitochondrial DNA sequences	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Smithsonian Institution,Department of Entomology,Washington,DC,20560**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Same as Report (SAR)**	**22**	

proposed by Sallum *et al.* (2000). A morphological cladistic analysis of Culicidae reconstructed *Chagasia* as the sister group of the clade consisting of *Anopheles* + *Bironella* (Harbach & Kitching, 1998). Molecular studies consistently support a close relationship between *Anopheles* and *Bironella*, but disagree on the details (Besansky & Fahey, 1997; Foley *et al.*, 1998). Sequence data for the mitochondrial ND5 gene and the combined ND5 and 28S rRNA D2 gene regions reconstructed *Bironella* as nested within a clade otherwise consisting of species of genus *Anopheles* (Krzywinski *et al.*, 2001a). Combined analyses of the ND5, D2, nuclear G_6pd and white genes similarly showed *Bironella* nested within *Anopheles* (Krzywinski *et al.*, 2001b). However, when the nuclear white gene data were analysed alone, *Bironella* was reconstructed as the sister genus to *Anopheles* (Besansky & Fahey, 1997; Krzywinski *et al.*, 2001b)

Based on four non-homoplastic morphological synapomorphies, Sallum *et al.* (2000) placed *Bironella* within genus *Anopheles*. There is only ambiguous morphological support for the precise relationship of *Bironella* with members of subgenera *Anopheles*, *Stethomyia* Theobald and *Lophopodomyia* Antunes.

Genus *Anopheles* formerly included six subgenera: *Anopheles*, the distribution of which is nearly world-wide; *Cellia* Theobald, which is restricted to the Old World tropics; and the four Neotropical subgenera *Lophopodomyia*, *Kerteszia* Theobald, *Nyssorhynchus* Blanchard and *Stethomyia* (Knight & Stone, 1977; Harbach, 1994). Morphological cladistic analysis found no support for the separate subgeneric status of either *Lophopodomyia* or *Stethomyia,* which were synonomized with subgenus *Anopheles* (Sallum *et al.*, 2000). Foley *et al.* (1998) suggested the paraphyly of *Cellia*; however, its monophyly was demonstrated by Sallum *et al.* (2000) and Krzywinski *et al.* (2001a,b). Previous systematic work by Foley *et al.* (1998) and Sallum *et al.* (2000) supported the paraphyly of subgenus *Anopheles*, whereas Krzywinski *et al.* (2001a,b) found support for the monophyly of the subgenus. Consequently, the systematics of subgenus *Anopheles* remain unresolved.

The goal of the present study was to use the genes COI and COII, a fragment of 18S rDNA and the expansion D2 region of 28S rDNA to test the current morphology-based phylogenetic hypothesis for Anophelinae (Sallum *et al.*, 2000). More specifically, this study aimed to test hypotheses of the non-monophyly of genus *Anopheles*, the paraphyly of subgenus *Anopheles* with respect to *Bironella* and the taxonomic status of each of subgenera *Anopheles*, *Stethomyia*, *Lophopodomyia* and genus *Bironella*.

The expansion D2 region of the nuclear large subunit 28S rDNA region was employed in this study because it appears to contain phylogenetic information for reconstructing relationships among insects at several appropriate taxonomic levels (Porter & Collins, 1996; Gimeno *et al.*, 1997; Belshaw *et al.*, 1998). Likewise, the 18S rDNA region, although evolving far too rapidly to reconstruct phylum- or subfamily-level relationships (Abouheif *et al.*, 1998; Maddison *et al.*, 1999), has proven useful for reconstructing a rapid radiation that occurred less than 40 Myr ago (Winnepenninckx *et al.*, 1998). The latter region has the additional advantage of having been well sampled in numerous animal clades, including arthropods (Giribet & Ribera, 1998, 2000). The mitochondrial cytochrome oxidase subunits I and II have proven useful as molecular markers for reconstructing evolutionary relationships among diverse arthropod genera and species (Simon *et al.*, 1994; Lunt *et al.*, 1996; Frati *et al.*, 1997; Wetterer *et al.*, 1998).

Materials and methods

Taxon sampling

The species sampled for this study, and the sources of specimens, are listed in Table 1. Thirty-two species of Anophelinae comprise the ingroup, consisting of *Bironella* (synonomized with genus *Anopheles* by Sallum *et al.*, 2000), *Chagasia* and *Anopheles*. For genus *Anopheles*, samples include members of all six subgenera recognized by Harbach (1994): *Anopheles*, *Cellia*, *Lophopodomyia*, *Kerteszia*, *Nyssorhynchus* and *Stethomyia*. Except for subgenera *Stethomyia* and *Lophopodomyia*, at least two species of each subgenus were sampled. Four species comprise the outgroup, *Aedeomyia* (*Aedeomyia*) *squamipennis*, *Uranotaenia* (*Uranotaenia*) *lowii*, *Ochlerotatus* (*Protomacleaya*) *triseriatus* and *Toxorhynchites* (*Toxorhynchites*) *amboinensis*. Portions of the mitochondrial genes cytochrome oxidase c subunit I (COI) and cytochrome oxidase c subunit II (COII), the nuclear 18S small subunit rRNA and the D2 variable region of the nuclear large subunit 28S rRNA were amplified and sequenced for one individual of each species. The region of each gene that was sequenced and the sequences and positions of the primers used in this study are listed in Table 2. GenBank numbers for all included taxa are AF417695–417730 for COI, AF417731–417766 for COII, AF417767–417801 for 18S and AF417802–417835 for D2 28S.

DNA extraction

Genomic DNA was extracted from specimens that were initially preserved by being frozen at −80 °C, initially preserved in 100% ethanol at ambient temperature in the field and subsequently frozen at −80 °C or preserved at room temperature in 70% ethanol. DNA was extracted from individual mosquitoes of each species following the methods described by Wilkerson *et al.* (1993). Whenever possible, male genitalia or larval and pupal exuviae, or all three, were mounted on slides and deposited in the National Museum of Natural History (NMNH), Smithsonian Institution and the Faculdade de Saúde Pública, Universidade de São Paulo (FSP-USP). DNA voucher specimens are deposited in the NMNH.

Table 1. Taxonomic list, stage and source of specimens used in this study.

Species	Stage	Locality of specimen
Ingroup		
An. aquasalis Curry	Female	Brazil, Rio de Janeiro, Magé
An. nuneztovari Gabaldón	Female	Brazil, Pará, Belém
An. marajoara Galvão & Damasceno	Female	Brazil, Mato Grosso, Peixoto de Azevedo
An. darlingi Root	Female	Brazil, Mato Grosso, Peixoto de Azevedo
An. albitarsis Lynch Arribalzaga	Female	Paraguay, Alto Parana, Hernadarias
An. rondoni (Neiva & Pinto)	Female	Brazil, Mato Grosso, Peixoto de Azevedo
An. albimanus Wiedmann	Female	Nicaragua, Zelaya Department, Caño Colorado
An. triannulatus (Neiva & Pinto)	Female	Brazil, Mato Grosso, Salobra
An. pseudopunctipennis Theobald	Female	Nicaragua, Zelaya, El Rama
An. coustani Laveran	Female	Kenya, locality not specified
An. puntimacula Dyar & Knab	Male	Nicaragua, Zelaya Department, Caño Colorado
An. eiseni Coquillett	Male	Brazil, São Paulo, Cananéia, Galiléia Reserve
An. intermedius (Peryassu)	Male	Brazil, Espirito Santo, Aguia Branca
An. freeborni Aitken	Male	U.S.A., NIH colony, locality not specified
An. atropos Dyar & Knab	Female	U.S.A., North Carolina, Topsail
An. punctipennis (Say)	Female	U.S.A., North Carolina, Abbotts Cr
An. judithae Zavortink	Female	U.S.A., Arizona, Patagonia National Park
An. dirus Peyton & Harrison	Female	Thailand, CDC colony, established by Esau & Scalon in 1963
An. farauti Laveran	Female	CDC colony, from US military colony
An. arabiensis Patton	Female	Kenya, Ahero, CDC colony, AHERO strain
An. gambiae Giles	Female	Gambia, CDC colony, G3 strain
An. stephensi Liston	Female	WRAIR colony
An. funestus Giles	Female	Kenya, locality not specified
An. minimus A Theobald	Female	Thailand (EK 461)
An. sundaicus (Rodenwaldt)	Female	Thailand (SUN 4)
An. subpictus Grassi	Female	Thailand (sb3)
An. bellator Dyar & Knab	Female	Brazil, São Paulo, Cananéia, Vilarinho
An. cruzii Dyar & Knab	Female	Brazil, São Paulo, Cananéia, Vilarinho
An. squamifemur Antunes	Male	Ecuador, Teputini Biodiversity Station
An. acanthotorhynus Komp	Male	Peru, Iquitos, Porto Almendras
Bi. gracilis (Theobald)	Female	Papua New Guinea, Sepikarca
Ch. bathana (Dyar)	Male	Ecuador, Teputini Biodiversity Station
Outgroup		
Ad. squamipennis Lynch Arribalzaga	Female	Brazil, São Paulo, Pariquera-Açu
Oc. triseriatus (Say)	Larva	U.S.A., Virginia, Alexandria
Ur. lowii Theobald	Female	Brazil, São Paulo, Pariquera-Açu
Tx. amboinensis (Doleschall)	Pupa	CDC, Puerto Rico colony

DNA amplification and sequencing

A fragment of approximately 903 bp of COI, representing 58.56% of the mitochondrial gene of *An. quadrimaculatus* Say (NC_000875), was amplified and sequenced, or sequenced, with the primers listed in Table 2.

COII fragments of approximately 605 bp, representing 88.18% of the gene in *An. quadrimaculatus*, were amplified and sequenced with the primers listed in Table 2. Thermal cycling conditions followed the protocol of Foley *et al.* (1998).

For the D2 variable expansion region of 28S rRNA, 499 bp (*Ch. bathana*) to 590 bp (*An. intermedius*) were amplified and sequenced using the primers listed in Table 2. The primers D2F and D2R are shortened versions of CP12 and CP15 of Porter & Collins (1996). *Anopheles dirus* and *An. funestus* were amplified using CP12 and CP15 and sequenced with D2F and D2R.

For the 18S small subunit rRNA, a fragment of 728–800 bp, representing 36.05–39.62% of the gene in *An. annulipes* A, was amplified and sequenced using primers whose design was based on the complete multiple sequence alignment of species from the *An. annulipes* complex (AF121053–AF121063). The sequence and position of each primer are listed in Table 2.

Standard protocols were used for all PCR amplifications (Palumbi, 1996). When amplifications gave poor results, hot-start PCR was used or PCR product was reamplified. Double-stranded PCR products were purified with QIAquick PCR purification kit (Qiagen, Operon, Alameda, CA, U.S.A.) or with PEG precipitation (20% polyethylene glycol (PEG) 8000/2.5 M NaCl) and sequenced directly using the primers listed in Table 2. All sequencing reactions were carried out using ABI Big Dye or ABI FS terminator chemistries (both PE Applied Biosystem, Foster City, CA,

Table 2. Sequences of COI, COII, 18S and D2 primers used in this study.

Designation	Sequence (5′ – 3′)	Use	Position	Reference
Fly5	TGTTTTAGCTGGAGCAATTACAAT	PCR/sequencing	601–624[3]	Lunt *et al.* (1996)
Fly10	AATGCACTAATCTGCCATATTAG	PCR/sequencing	tRNAleu	Lunt *et al.* (1996)
COIFg	AGTATTAGCAGGAGCTATTACTAT	PCR/sequencing	594–617[3]	This study
Fly5IP[1]	GGATTATTAGGATTTATTGT	Sequencing	842–861[3]	This study
Fly10IP[1]	GCAAATAATGAAATTGTTCT	Sequencing	1373–1392[3]	This study
COIF2AS[1]	GCTCATTTTCATTATGT	Sequencing	1124–1140[3]	This study
COIR2AS[1]	GAAGTAAAATAAGCTCG	Sequencing	905–921[3]	This study
COIF2g[1]	GGATTTATTGTTTGAGCTCA	Sequencing	851–870[3]	This study
COIR2g[1]	CGTCGAGGTATTCCGGCTAA	Sequencing	1298–1317[3]	This study
COIIF1	ATGGCAACATGAGCAAATT	PCR/sequencing	8–26[3]	This study
COIIR1	ATTCTTTCAATTACAATTGG	PCR/sequencing	629–648[3]	This study
COIIR2	AATTTATAGGAATTCTTTC	PCR/sequencing	641–659[3]	This study
18SF1	AGCTCCACTAGCGTATATTAAAGT	PCR/sequencing	586–609[4]	This study
18SR1	TTAACCAGACAAATCGATCCACGA	PCR/sequencing	1446–1469[4]	This study
D2F[2]	AGTCGTGTTGCTTGATAGTG	PCR/sequencing	288–307[5]	This study
D2R[2]	CTTGGTCCGTGTTTCAAGAC	PCR/sequencing	821–840[5]	This study
CP12	GTGGATCCAGTCGTGTTGCTTGATAGTGCAG	PCR	280–310[5]	Porter & Collins (1996)
CP15	GTGAATTCTTGGTCCGTGTTTCAAGACGGG	PCR	818–847[5]	Porter & Collins (1996)

[1]Internal primers used in sequencing reactions.
[2]Shortened version of CP12 and CP15.
[3]Nucleotide position relative to COI and COII sequence of *An. quadrimaculatus* (NC_000875), except for Fly10, whose position is relative to COI gene in *D. yakuba*.
[4]Nucleotide position relative to 18S sequence of *An. annulipes* A (AF121053).
[5]Nucleotide position relative to 28S sequence of *An. albimanus*.

U.S.A.). Templates were sequenced in both directions and sequences were generated with an ABI 373 automated sequencer (PE Applied Biosystem).

Sequence alignment

Nucleotide sequences of COI and COII were aligned using the multiple alignment program Clustal W 1.6 (Thompson *et al.*, 1996), adjusted by visual inspection and guided by comparison with the complete published sequence of *An. quadrimaculatus*, using both nucleotides and amino acids. The DNA sequence of *An. quadrimaculatus* was translated into amino acids based on the invertebrate genetic code using MacClade version 4.0 PPC (Maddison & Maddison, 2000). Nuclear 18S sequences and D2 28S rRNA sequences were automatically aligned using Clustal X 1.8 (Thompson *et al.*, 1997). The alignment of 18S sequences was guided by the primary and secondary structures of *An. pseudopunctipennis*, which were downloaded from rRNA database (http://rrna.uia.ac.be) (De Peer *et al.*, 1997). Aligning of 18S was done under 'profile alignment' using default conditions. Profile 1 was the sequence downloaded from the rRNA database and profile 2 consisted of the sequences generated by the present study. The alignment of D2 sequences was carried out using multiple alignment mode under gap opening and extension, as follows: GO50/GE10, GO20/GE10, GO10/GE1 and GO10/GE5. The GO10/GE5 alignment was used in the analyses. The 18S and D2 sequence alignments were adjusted by visual inspection using MacClade version 4.0 PPC (Maddison & Maddison, 2000). To utilize shared, multiple-nucleotide insertion/deletion events ('indels') as characters in phylogenetic analyses conducted under the parsimony criterion, multiple-site deletions were scored based on the method of 'simple indel coding' (Simmons & Ochoterena, 2000). Only those regions that could be scored unambiguously were coded for gaps and included in parsimony analyses. Three hyper-variable regions consisting of seventy-two nucleotide positions (10.54% of the positions in the D2 gene region) were found to be unalignable and were excluded from all analyses (including indel coding). Sequence data for all genes were combined into a single data matrix using MacClade version 4.0 PPC.

Phylogenetic analysis

The data consist of 1596 bp of nuclear ribosomal DNA (rDNA) sequence (913 bp from the 18S subunit and 683 bp from the D2 region of the 28S subunit) and 1507 bp of mitochondrial DNA (mtDNA) sequences (903 bp from COI and 604 bp from COII). Parsimony (MP) and maximum likelihood (ML) analyses were carried out on various datasets, including 18S only, 28S only, 18S and 28S (i.e. rDNA only), mtDNA nucleotides (i.e. COI and COII), mtDNA, translated to amino acid sequence, and mtDNA and rDNA. It was not possible to obtain 28S sequence for *An. acanthotorynus* (representing subgenus *Stethomyia*) and *Bi. gracilis*; sequence for the 18S rDNA gene region could not be obtained for *An. farauti*. Because it was suspected that the high concentration of missing data in *An. acantho-*

torynus may have artificially affected the results of some analyses (Nixon & Wheeler, 1992; Wiens & Reeder, 1995; Wiens, 1998; Klompen *et al.*, 2000), most analyses were conducted with *An. acanthotorynus* both included and excluded. Although there were missing data for *An. farauti* and *Bi. gracilis* as described above, these species consistently grouped with their sister taxa (*An. dirus* and *An. squamifemur*, respectively), with strong branch support in all analyses. Thus, missing data were not considered to be a problem in these species and analyses, and then exclusion was pursued.

Parsimony analyses. Parsimony analyses were implemented in PAUP 4.0b4a (Swofford, 1998) using the heuristic search option with TBR branch-swapping and with parsimony-uninformative characters excluded. To ensure that multiple 'islands' of most parsimonious trees were identified (Maddison, 1991), 500 random-taxon-addition replicate analyses were carried out for the unweighted analyses and 200 random-taxon-addition replicate analyses were carried out for the successive approximations weighted analyses. For the successive approximations weighted analyses, character weights were based on the maximum value of the rescaled consistency index and iterative rounds were continued until character weights stabilized (Farris, 1969; Carpenter, 1988). Bootstrapping (Felsenstein, 1985) under parsimony utilized 1000 pseudoreplicates, with ten random-taxon-addition replicates per pseudoreplicate; parsimony-uninformative characters were excluded.

Maximum-likelihood (ML) analyses. To obtain an appropriate substitution model and model parameter values, as well as an optimal starting tree for branch-swapping under ML, one or more optimal trees obtained by parsimony analysis were evaluated under 56 'models' of evolution using the computer program ModelTest 3.0 (Posada & Crandall, 1998), which compares fourteen basic substitution models. All fourteen models were evaluated with and without rate heterogeneity. Rate heterogeneity was accommodated in three ways: using a gamma model with six rate categories, using an invariant sites model and using a gamma plus invariant sites model (Swofford *et al.*, 1996). Using a standard likelihood ratio test, the likelihood scores of each of the parsimony trees were compared across nested models with the computer program ModelTest 3.0 (Posada & Crandall, 1998). In pairwise comparisons in which the improvement in likelihood imparted by a more complex model was not found to be significant, the simpler model was chosen. Likelihood scores were also compared across both nested and non-nested models in ModelTest 3.0 using the Akaike Information Criterion (AIC). In the three cases in which the tests disagreed about model choice, analyses were conducted using the simpler model as discussed below. All datasets were additionally analysed using the most complex model available, GTR + I + Γ.

Employing the adopted model and using the optimal parsimony tree (either the single tree chosen by unweighted parsimony or successive approximations weighted parsimony or, if multiple equally optimal parsimony trees were identified, the tree from this group with the highest likelihood score under the adopted model) as the starting tree for branch-swapping, five iterative rounds of maximum-likelihood analyses were carried out, proceeding from those using less intensive to those using more intensive branch-swapping regimens. The most likely tree identified during each of the first four ML search rounds was used as the starting tree for the next search round, both for the calculation of updated parameter values and for the initiation of branch-swapping. Branch-swapping regimens in the five rounds were, respectively, nearest-neighbour interchange (NNI), subtree pruning-regrafting (SPR), SPR, tree bisection-reconnection (TBR) and TBR. In all rounds except round 4, the Rogers–Swofford approximation limit was set to 0.05 ('approxlim = 5') and all optimal trees were saved during swapping. In round 4, the Rogers–Swofford approximation limit was set to 0.02 ('approxlim = 2') and only one optimal tree was saved during swapping ('mulpars = no'). Bootstrapping (Felsenstein, 1985) under the ML criterion utilized 100 pseudoreplicates, with a single random-taxon-addition starting tree per pseudoreplicate and TBR branch-swapping. To shorten the ML bootstrapping tree-search times, two compromises with regard to search thoroughness were made: the Rogers–Swofford approximation limit was set to 0 (i.e. no branch-length optimizations were pursued beyond the parsimony-based branch-length estimates calculated initially by PAUP) and ML model parameter values were set to the optimal (i.e. final) values estimated during the likelihood search procedure described above.

Statistical tests. The Shimodaira–Hasegawa tests were performed using RELL approximation as described in Goldman *et al.* (2000), which is based on Shimodaira & Hasegawa (1999). The likelihood ratio test for monophyly was performed as in Huelsenbeck *et al.* (1996).

Bayesian analyses used MrBayes version 1.0 (Huelsenbeck & Ronquist, 2000). For the Bayesian analysis a GTR + Γ model was used, with six categories of rates, and with estimated base frequencies. Program default values for the prior probabilities were used. The MCMC was allowed to run 200 000 generations, and sampled every 100 generations after a burn-in of 100 000 generations. The log likelihood output of the chain was plotted to confirm that the chain had stabilized during the burn-in.

Results

Sequence characteristics

The alignment of 18S rRNA sequences produced sequences 913 bp in length, of which 378 sites (41.4%) were variable, 270 (29.6%) were parsimony informative and 276 included insertion/deletion (indel) events. The 683 bp of 28S included 327 sites that required indels in some species; 72 bp of unalignable 'hypervariable' regions

were entirely excluded from all analyses. These 'hypervariable' regions are positions 223–239, 321–344 and 448–478. The remaining 611 bp contained 432 variable sites (70.7%), of which 340 (55.6%) were parsimony informative. For parsimony analyses, some indel regions in both the 18S and 28S sequences were coded as character data following the conservative simple indel coding method of Simmons & Ochoterena (2000), as described below, producing an additional 314 binary (present/absent) characters (164 for D2 and 150 for 18S) that were used in the parsimony analyses.

Alignment of the mitochondrial genes was straightforward, as no indels were required. The alignment of COI mtDNA sequences produced sequences 903 bp in length. Of these, 361 sites were variable (39.9%) and 289 (32.0%) were parsimony informative. The 604 bp of COII sequences contained 278 variable sites (46.0%), of which 217 (35.9%) were parsimony informative. For amino acid analyses, the mtDNA data were translated into a sequence of 502 amino acids using the *Drosophila* genetic code in MacClade 4.0. Alignments are available from TRS, PGF and MAMS upon request.

The method used for coding indels is the conservative simple indel coding method. Briefly, all multiple-site base deletions that had different 5′ or 3′, or both, termini were scored as separate present/absent characters, and multiple-site deletions representing subsets of longer, completely overlapping deletions were coded as inapplicable for the indel character being scored (Simmons & Ochoterena, 2000).

For the 18S rDNA gene region, uncorrected ('p') sequence distances ranged from 0 (between some species within subgenus *Nyssorhynchus*) to 0.19025 (between *An. judithae* and *An. minimus* A); for the 28S rDNA region, distances ranged from 0.00198 (again between some species within subgenus *Nyssorhynchus*) to 0.34545 (between *Ad. squamipennis* and *An. coustani*). For the COI mtDNA region, distances ranged from 0.00221 (between *An. gambiae* and *An. arabiensis*) to 0.17702 (between *Ad. squamipennis* and *An. arabiensis*), and for the COII mtDNA region from 0.00331 (also between *An. gambiae* and *An. arabiensis*) to 0.21523 (between *Ad. squamipennis* and *An. punctipennis*). Nucleotide frequencies for all four regions are summarized in Table 3. The COI and COII regions showed considerable A + T nucleotide bias (Table 3), which is consistent with other insect mitochondrial genes (Crozier & Crozier, 1993; Simon *et al.*, 1994; Frati *et al.*, 1997; Chippindale *et al.*, 1999). Standard χ^2 tests for base homogeneity implemented in PAUP 4.0b4a were unable to reject homogeneity of base frequencies among species sequences for any of the four genes when all sites were included ($P > 0.99$). When sites observed to be non-variable were excluded, homogeneity could not be rejected ($P > 0.20$) for the 18S, 28S and COI datasets; however, homogeneity was rejected for COII ($P = 0.032$), indicating possible non-stationarity of base substitution processes across the phylogeny relating these sequences.

To test the congruence of the separate 18S, 28S, COI and COII datasets, we employed the Incongruence Length Difference test of Farris *et al.* (1995), implemented as the 'Partition Homogeneity Test' in PAUP 4.0b4a with 1000 replicates and ten random-taxon-addition tree searches per replicate and with invariant sites excluded (Cunningham, 1997). The results (Table 4) indicate that, by the criterion of this test, congruence cannot be rejected for the COI and COII datasets ($P = 0.932$); however, the 18S and 28S datasets are incongruent with each other as well as with the mtDNA data, and this incongruence persists whether or not the 'indel-coded' characters are included ($P = 0.001$ in all comparisons). Some authors have suggested that datasets suspected to be incongruent should not be combined, whereas other authors disagree (Bull *et al.*, 1993; Eernisse & Kluge, 1993; de Queiroz *et al.*, 1995; Nixon & Carpenter, 1996; DeSalle & Brower, 1997). In this study, we have chosen to analyse the incongruent datasets both individually as well as in all possible combinations. This approach localizes areas of dataset agreement on topologies. In general, when datasets disagree about the monophyly of a clade, branch support for that clade is reduced when the datasets are combined. Alternatively, in cases where the datasets agree, support for the clade increases on the topology inferred from the combined data (Thornton & DeSalle, 2000). Our approach avoids demonstrated problems with using the ILD test as the ultimate arbiter of data combinability, including the study of Mitchell *et al.* (2000), which demonstrated that when datasets found to be incongruent by the ILD test (Farris *et al.*, 1995) were combined, overall support was greater on the combined topology than on the topologies produced by either dataset analysed alone; and the study of Yoder *et al.* (2001), which demonstrated that phylogenetic accuracy can increase by combining datasets found to be incongruent by the ILD test.

Table 3. Mean nucleotide frequencies for the genes used in this study.

Gene	A	C	G	T
18S rDNA	0.26786	0.21599	0.27235	0.24380
28S rDNA	0.23327	0.27022	0.29604	0.20047
COI mtDNA	0.31276	0.15556	0.14268	0.38901
COII mtDNA	0.35277	0.14313	0.11785	0.38625

Table 4. ILD test results.

ILD tests	*P* values
18S vs 28S (with indels)	0.001
18S vs 28S (without indels)	0.001
COI vs COII	0.932
rDNA vs mtDNA (with indels)	0.001
rDNA vs mtDNA (without indels)	0.001
18S vs mtDNA (with indels)	0.001
18S vs mtDNA (without indels)	0.001
28S vs mtDNA (with indels)	0.001
28S vs mtDNA (without indels)	0.001

Phylogenetic analysis of the nuclear rDNA gene regions

Parsimony analyses of the rDNA datasets included the indel information coded as described above. Aside from the use of the resulting optimal parsimony trees as the starting points for ML analyses, this indel information played no role in ML analyses.

18S parsimony. Parsimony analysis of the 18S dataset (including the indel-coded characters) identified two most parsimonious trees (MPTs) of length = 1247, CI = 0.496 and RI = 0.704. Analysis using successive approximations character weighting identified the same two trees.

18S likelihood. One of the SWTs was evaluated in the program ModelTest 3.0. The likelihood ratio test found the TrNef + I + Γ model, but with equal base frequencies, a proportion of sites invariant and gamma-distributed rates to be significantly better fitting than the next less complex model ($P = 0.000067$).

Complete likelihood analyses were subsequently conducted using both this model and the most complex model, GTR + I + Γ (Rodríguez *et al.*, 1990; general time reversible with a proportion of sites invariant and gamma-distributed rates). The TrNef + I + Γ analysis found two equally likely trees (differing only in the placement of *An. triannulatus* within *Nyssorhynchus*) with a log likelihood of −6364.71090. Analysis using GTR + I + Γ yielded a single tree, identical with one of the two trees found under the simpler model, with log likelihood of −6363.49490.

Bootstrap analyses. For the 18S data, bootstrap support for most clades is moderate to strong in the MP analysis. In contrast, only thirteen clades achieved >50% bootstrap support in the ML analyses. The most basal relationships are weakly supported in both ML and MP analyses (< 50% bootstrap proportions), but they are better supported in MP bootstrap analyses.

The 18S MP tree confirms the monophyly of Anophelinae, the basal position of *Chagasia* within the subfamily and the grouping *Anopheles* plus *Bironella*. Because *An. judithae* does not group with subgenus *Anopheles* in either the MPTs or with a >50% bootstrap proportion, the monophyly of subgenus *Anopheles* is not supported. *Stethomyia* is reconstructed as the sister group of *Cellia* with strong support, and the monophyly of the latter is also strongly supported (Table 5), but deeper relationships within *Cellia* are generally unresolved. The monophyly of the clade consisting of *Kerteszia*, *Lophopodomyia*, *Nyssorhynchus* and *Bironella* is strongly supported. Three major clades are recovered within this larger clade: a basal clade containing members of subgenus *Kerteszia*, the clade *Bironella* + *Lophopodomyia*, and a *Nyssorhynchus* clade. A sister-group relationship between *Bironella*, *Lophopodomyia* and *Nyssorhynchus* is strongly supported (Table 5), but basal relationships between major groups within genus *Anopheles* (including *Bironella*) are not well resolved.

Under ML, the monophyly of Anophelinae (including *Chagasia*) is poorly supported (< 50% bootstrap proportion), whereas the monophyly of the clade consisting of *Anopheles* + *Bironella* is strongly supported (Table 5). Resolution within that clade is generally poor. Exceptions include the monophyly of *Kerteszia* and *Cellia*. A sister-group relationship between *Lophopodomyia* and *Bironella* is moderately well supported (Table 5).

28S parsimony. Parsimony analysis of the 28S dataset identified sixty-eight MPTs of parsimony-informative length = 1477, CI = 0.470 and RI = 0.668. Analysis using successive approximations character weighting identified nine trees, a subset of the MPTs.

Table 5. Bootstrap support for relationships of and within genus *Anopheles*.

	18S MP/ML	28S MP/ML	18S + 28S MP/ML	mtDNA (n) MP/ML	mtDNA (aa) MP	rDNA + mtDNA MP/ML
Sister group						
Chagasia and *Anopeles*	88/< 50	99/96	100*,100/96*,96	92/100	99	100*,100/100*,100
Monophyly						
Anopheles (including *Bironella*)	100/100	NA/NA	100*,100/89*,100	< 50/< 50	< 50	85*,100/100*,100
Subgenera						
Anopheles	< 50/< 50	84/58	81*,84/77*,75	< 50/< 50	< 50/< 50	< 50*,74/70*,69
Kerteszia	100/99	C/C	100*,100/97*,100	100/100	100/100	100*,100/100*,100
Cellia	100/97	100/99	100*,100/91*100	< 50/< 50	< 50/< 50	98*,100/97*,100
Nyssorhynchus	82/< 50	60/90	99*,98/< 50*,C	< 50/< 50	< 50/< 50	99*,100/59*,79
Other groupings						
(*Lophopodomyia*, *Bironella*)	100/77	NA/NA	99*,100/75*,92	< 50/< 50	< 50/< 50	96*,94/85*,87
(*Stethomyia*, *Cellia*)	90/< 50	NA/NA	94*,NA/< 50*,NA	< 50/< 50	< 50/< 50	< 50*,NA/C*,NA
(*Kerteszia*, *Nyssorhynchus*)	C/C	100/90	99*,99/67*,81	< 50/< 50	< 50/< 50	90*,98/< 50*,95
BLNK	92/< 50	C/C	53*,62/< 50*,69	< 50/< 50	< 50/< 50	51*,63/C*,81
Anopheles, BLNK	< 50/< 50	C/C	71*,78/< 50*,50	< 50/< 50	< 50/< 50	< 50*,50/C*,C

C, contradicted by an alternative relationship appearing on bootstrap tree. NA = not applicable.
*, *Stethomyia* included, *Stethomyia* excluded. BLNK = *Bironella*, *Lophopodomyia*, *Nyssorhynchus*, *Kerteszia*.

28S likelihood. One of the nine SWTs was evaluated in the program ModelTest 3.0. Both the likelihood ratio test ($P < 0.000001$) and the Akaike information criterion found the TrN + I + Γ model, but with unequal base frequencies, a proportion of sites invariant and gamma-distributed rates to be significantly better fitting than the next less complex model. Complete likelihood analyses were subsequently conducted using both this model and the most complex model, GTR + I + Γ. The TrN + I + Γ analysis recovered a single most likely tree with a log likelihood of −6535.24527. Analysis using GTR + I + Γ yielded a single tree identical with the tree found using the simpler model, with log likelihood of −6533.28831.

Bootstrap analyses. The MP bootstrap support for the 28S data is moderate to strong for most clades; ML bootstrap support values, however, are generally weak. The most basal relationships are generally better supported in MP analyses than in ML analyses. The monophyly of Anophelinae is well supported in both MP and ML analyses (Table 5). Due to missing sequence data for both *Bironella* and *Stethomyia*, however, the 28S dataset is unable to resolve questions about the relationships of these groups within Anophelinae. Unexpectedly, MP analysis places *Lophopodomyia* in the most basal position within genus *Anopheles*, and *Cellia* as the sister group to the clade formed by subgenera *Anopheles*, *Kerteszia* and *Nyssorhynchus*. The monophyly of *Cellia* is strongly supported and that of subgenus *Anopheles* is moderately well supported. A sister-group relationship between *Kerteszia* and *Nyssorhynchus* is strongly supported. In contrast, the MP tree does not recover the monophyly of *Kerteszia*, and support for the monophyly of *Nyssorhynchus* is low (Table 5).

In the optimal ML tree, *Lophopodomyia* is basal within genus *Anopheles* and a clade consisting of *Kerteszia* and *Nyssorhynchus* is the sister group to a clade consisting of *Cellia* and *Anopheles*. However, these basal relationships are supported by <50% bootstrap proportions. The monophyly of *Cellia* is strongly supported; however, basal relationships within this clade are not well resolved. The monophyly of subgenus *Anopheles* is weakly supported, and the paraphyly of *Kerteszia* with respect to *Nyssorhynchus* is relatively well supported (Table 5).

Combined 18S and 28S parsimony. Parsimony analysis of the combined 18S and 28S (rDNA) datasets produced forty-one MPTs with parsimony-informative length = 2764, CI = 0.475 and RI = 0.676. Analysis using successive approximations character weighting identified a subset of the MPTs consisting of twenty-two trees. Parsimony analysis of the combined 18S and 28S (rDNA) datasets from which *Stethomyia* (*An. acanthotorynus*) was excluded produced eighty-two MPTs with parsimony-informative length = 2684, CI = 0.485 and RI = 0.684. Except for decreased resolution within clade *Nyssorhynchus*, the strict consensus of these MPTs is entirely congruent with the strict consensus produced by the analysis in which *Stethomyia* (*An. acanthotorynus*) was included, and overall support for most clades increases. Analysis using successive approximations character weighting identified a subset of the MPTs consisting of twenty-two trees. Except for the position of *An. acanthotorynus*, these are the same twenty-two trees that were found in the analysis in which *An. acanthotorynus* was included (Fig. 1).

Combined 18S and 28S likelihood. One of the twenty-two SWTs was evaluated in the program ModelTest 3.0. The likelihood ratio test found the TrNef + I + Γ model to be significantly better fitting than the next less complex model ($P < 0.000001$), whereas the Akaike information criterion found the TIM + I + Γ model (limiting transversions to two or four rates) to be the best fitting. Because the model chosen by the likelihood ratio test is simpler, we chose to conduct complete likelihood analyses using both this simplest model (TrNef + I + Γ) and the most complex model available, GTR + I + Γ. The TrNef + I + Γ analysis yielded a single most likely tree with a log likelihood of −13168.63111. Analysis using GTR + I + Γ yielded a single tree, identical with the tree found using the simpler model, with a log likelihood of −13149.31266 (Fig. 2).

An identical procedure was followed in the likelihood analysis of the combined 18S and 28S dataset from which *Stethomyia* was excluded. When one of the twenty-two SWTs was evaluated for fifty-six models of evolution, the likelihood ratio test again favoured the TrNef + I + Γ model ($P < 0.000001$), but the Akaike information criterion now favoured the TrN + I + Γ. Again, two analyses were run, one using the simplest suggested model, TrNef + I + Γ and the other using the most complex available model, GTR + I + Γ. The former identified a single tree with a log likelihood of −12790.34050; the latter produced a tree with identical topology and a log likelihood of −12774.81865 (Fig. 2).

Significantly, except for the absence of *An. acanthotorynus*, the tree found in the ML analysis of the combined 18S and 28S dataset with *An. acanthotorynus* excluded is identical to the tree found by the ML analysis in which *An. acanthotorynus* was included.

Bootstrap analyses. Bootstrap support for most clades is moderate to strong in the MP analysis in which *An. acanthotorynus* is included. Relative to these results, overall support for most clades increased in the analysis from which *An. acanthotorynus* was excluded (Fig. 1). Bootstrap support for most clades is similarly strong in ML analyses, and, again, support values generally increase when *An. acanthotorynus* is excluded from ML analyses (Fig. 2).

The identical ML trees recovered under both the TrNef + I + Γ and the GTR + I + Γ models and with *An. acanthotorynus* either included or excluded from the analyses confirm the monophyly of Anophelinae and define three major clades within the clade consisting of genera *Anopheles* and *Bironella* (Fig. 2): a clade containing species of *Nyssorhynchus*, *Kerteszia* and *Lophopodomyia*, and genus *Bironella*; a clade corresponding to *Cellia*; and a clade corresponding to subgenus *Anopheles*. The *Cellia*

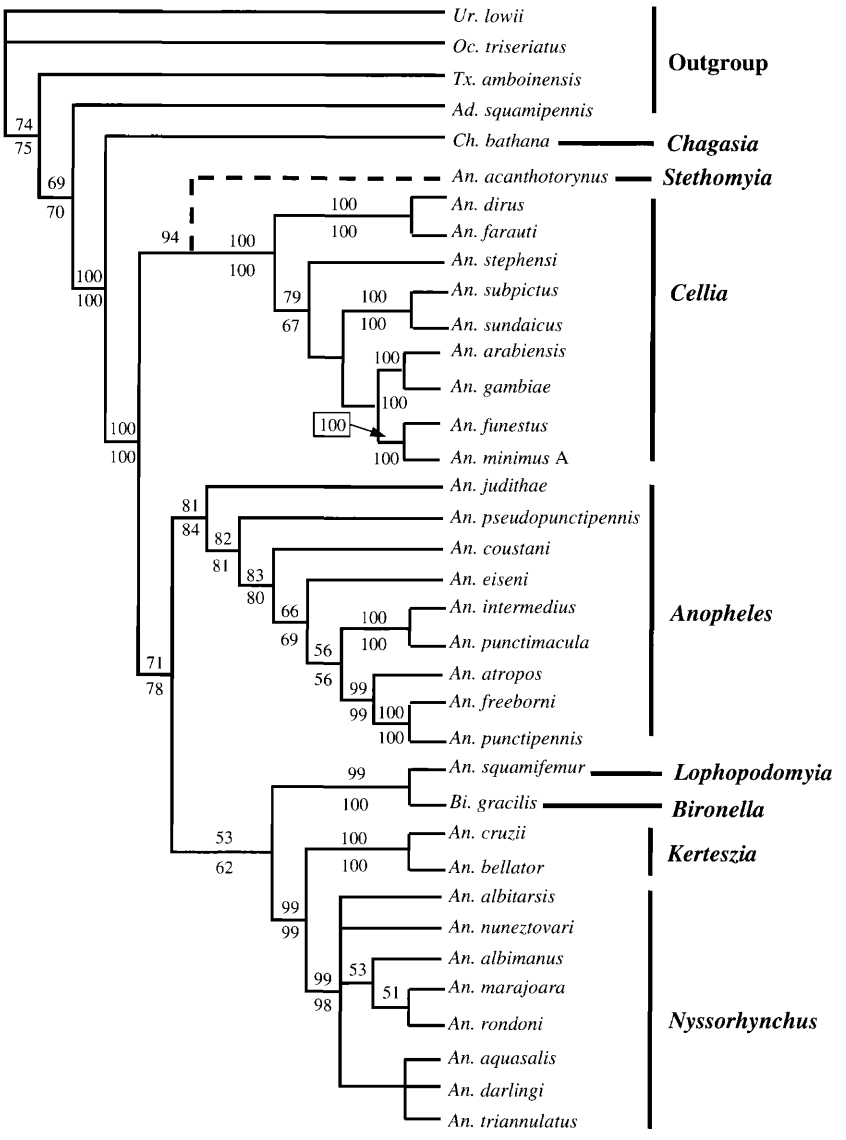

Fig. 1. Strict consensus of twenty-two equally parsimonious trees identified by successive approximations character weighting parsimony analyses of the combined 18S and 28S rDNA data with *An.* (*Stethomyia*) *acanthotorynus* excluded. Except for increased resolution within clade *Nyssorhynchus*, the strict consensus of these MPTs is entirely congruent with the strict consensus produced by the analysis in which *Stethomyia* was included. Numbers above and below branches indicate MP bootstrap proportions obtained when *Stethomyia* was included and excluded from the analyses, respectively. Dashed line indicates the position of *Stethomyia*. Forty-one equally parsimonious trees were identified by unweighted parsimony analyses; this subset of twenty-two trees was favoured by successive approximations character weighting analyses.

and *Anopheles* clades are sister groups, and the larger clade consisting of these groups is the sister group of the clade (*Nyssorhynchus* + *Kerteszia* + *Lophopodomyia* + *Bironella*). However, ML bootstrap analyses indicate that the combined 18S and 28S data provide only weak support for these relationships between the three major clades (Table 5). The monophyly of the major clade *Nyssorhynchus* + *Kerteszia* + *Lophopodomyia* + *Bironella* is moderately supported. Within the *Nyssorhynchus* + *Kerteszia* + *Lophopodomyia* + *Bironella* clade, a sister-group relationship between *Lophopodomyia* and *Bironella* is well supported (92% bootstrap proportion), the monophyly of a *Nyssorhynchus* + *Kerteszia* clade somewhat less so. *Kerteszia* is recovered as a monophyletic group nested within *Nyssorhynchus* as the sister group of *An. triannulatus*. However, this relationship is only weakly supported, as are relationships within *Nyssorhynchus* (Table 5).

The analysis favours the hypothesis of monophyly of subgenus *Anopheles*. However, bootstrap support for this conclusion is relatively low. Within subgenus *Anopheles*, the monophyly of the Arribalzagia Series is strongly supported (100% bootstrap proportion), as is the monophyly of the

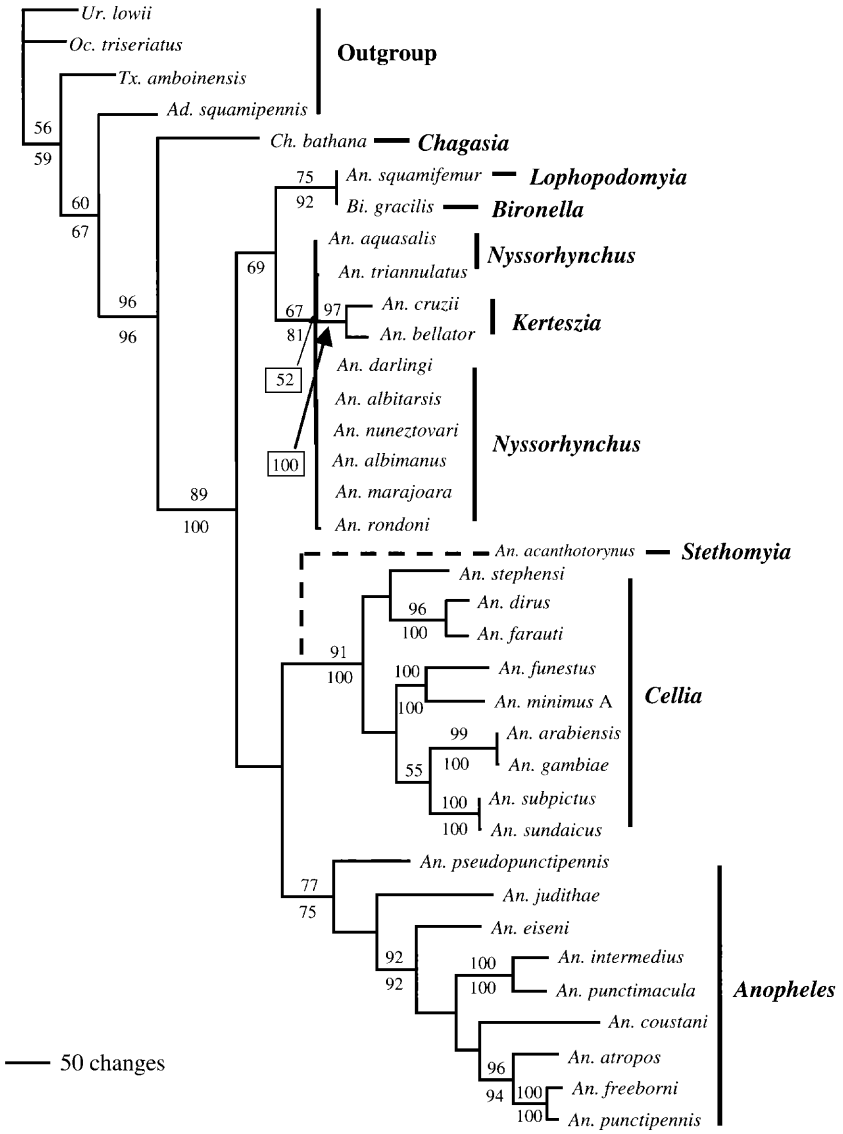

Fig. 2. The single tree identified by maximum likelihood analyses of the combined 18S and 28S rDNA data under both the TrNef + I + Γ and GTR + I + Γ models of nucleotide substitution in analyses in which *An.* (*Stethomyia*) *acanthotorynus* was both included and excluded. Numbers above and below branches indicate ML bootstrap proportions obtained when *Stethomyia* was included and excluded from the analyses, respectively. Dashed line indicates the position of *Stethomyia*.

group consisting of *An. atropos*, *An. freeborni* and *An. punctipennis* of the Anopheles Series, which otherwise appears to be polyphyletic. The monophyly of *Cellia* is strongly supported (Table 5). *Cellia* contains three major lineages: a clade consisting of two sister groups, one composed of species of the Neomyzomyia Series and one containing *An. stephensi* of the Neocellia Series; the Myzomyia Series, which is monophyletic; and the sister clade to the Myzomyia Series containing the Pyretophorus Series. However, bootstrap support for these relationships within *Cellia* is low and the monophyly of the Pyretophorus Series is supported by <50% bootstrap proportion. The inclusion of *Stethomyia* in analyses did not change ingroup topology, but bootstrap support for most clades decreased (Fig. 2, Table 5). An interesting exception to this rule is the monophyly of the Pyretophorus Series within *Cellia*, for which bootstrap support increased when *An. acanthotorynus* was included. In analyses in which *An. acanthotorynus* was included, the support for the placement of the clade *Bironella* + *Lophopodomyia* within the clade *Nyssorhynchus* + *Kerteszia* decreased. Although the bootstrap support is low (Table 5), it is worth noting that *Stethomyia* was reconstructed as the sister group to *Cellia* (Fig. 2).

The MP tree generated from the analysis with *Stethomyia* included is similar to that with *Stethomyia* excluded except for the lower bootstrap support value for phylogenetic

relationships within the clade *Nyssorhynchus* (Fig. 1). In addition, *Cellia* is recovered as the sister group to a monophyletic clade composed of *Anopheles* + *Lophopodomyia* + *Bironella* + *Kerteszia* + *Nyssorhynchus*. A comparison of the bootstrap proportions indicates that major disagreement between optimal MP and ML trees for the combined rDNA dataset is restricted to basal placement of the clade formed by *Cellia* + *Stethomyia* in the MP tree (Fig. 1), in contrast to the basal position of the clade *Nyssorhynchus* + *Kerteszia* + *Lophopodomyia* + *Bironella* in the ML tree (Fig. 2). Consequently, in the former tree the group *Nyssorhynchus* + *Kerteszia* + *Lophopodomyia* + *Bironella* was recovered as the sister group of *Anopheles*, but in the ML tree (*Cellia* + *Stethomyia*) is sister to *Anopheles*. Additionally, *Kerteszia* is nested within *Nyssorhynchus* in the ML topology, but MP reconstructed the separate monophyly of *Kerteszia*. Most clades are strongly supported by bootstrap proportions in the MP analysis (Table 5). The major differences between the results of the ML and MP topologies when *An. acanthotorynus* is excluded are, in general, the same as those found when *An. acanthotorynus* is included.

Phylogenetic analysis of the mitochondrial gene regions

Because the partition homogeneity test indicated no significant incongruence between the COI and COII datasets (Table 4), combing the two datasets for analysis is uncontroversial.

Parsimony. Parsimony analysis of nucleotide sequences for the combined (mtDNA) dataset identified two MPTs with parsimony-informative length = 3058, CI = 0.257 and RI = 0.356. Analysis using successive approximations character weighting identified a single SWT with parsimony-informative length = 3063, CI = 0.257 and RI = 0.354.

In addition to the analysis of nucleotide sequence, the mtDNA data were converted to amino acid sequences and analysed under the criterion of 'protein parsimony' (Felsenstein, 1996) using a step matrix constructed in MacClade 4.0. This analysis produced 37 MPTs with weighted length = 290, CI = 0.497 and RI = 0.620. Analysis using successive approximations character weighting identified a subset of eight of the MPTs.

Likelihood. One of the two SWTs from the MP nucleotide analysis was evaluated in the program ModelTest 3.0. Both the likelihood ratio test and the Akaike information criterion found the GTR + I + Γ model to be significantly better fitting than the next less complex model ($P < 0.000001$). Likelihood analysis using GTR + I + Γ yielded a single most likely tree with a log likelihood of −14976.15202.

Bootstrap analyses. Both MP and ML bootstrap support values for the mtDNA data are generally very poor, with only twelve clades achieving >50% bootstrap proportions in the MP analysis (Fig. 3) and thirteen in the ML analysis (Fig. 4). Aside from strong support for the monophyly of Anophelinae (92% under MP; 100% under ML), all of the relationships supported at >50% are recent, grouping at most five species. Under MP, only groupings consisting of two species (of which there are eight) achieve proportions of 70% or greater; groupings of three or more species (of which there are three, not counting Anophelinae) are supported by low proportions (57, 63 and 63%). The results of protein parsimony bootstrap analysis of mtDNA amino acid sequences are similar.

Similar patterns are obtained in the ML bootstrap analysis. Clades consisting of two species (of which there are six) are supported by >85% bootstrap proportions, whereas clades of three to five species (of which there are four) are supported by generally lower proportions (51, 63, 68 and 83%). The results of the MP and ML bootstrap analyses are consistent with the conclusion that the mtDNA regions analysed are far too rapidly evolving to recover the phylogenetic relationships of any but the most recently derived species groups and subgenera within Anophelinae. The fact that relationships between *Nyssorhynchus* species are among the only relationships recovered by bootstrap analyses of the mtDNA data, but not by the nuclear datasets, suggests that these divergences may have occurred too recently to be tracked by information in the ribosomal DNA data. This conclusion is also supported by the low genetic distances separating these species, detailed above.

Phylogenetic analysis of the combined nuclear and mitochondrial gene regions

Parsimony. Parsimony analysis of the combined rDNA (including indel-coded characters) and mtDNA datasets generated eight MPTs, length = 5915, CI = 0.355 and RI = 0.523. Analysis using successive approximations character weighting identified a single SWT, not one of the MPTs, with parsimony-informative length = 5916, CI = 0.355 and RI = 0.523. The SWT is similar to the tree obtained from an analysis of the rDNA data alone except for the position of *An. acanthotorynus*, which arises within subgenus *Anopheles* as the sister species of *An. judithae*, and the relationships within *Nyssorhynchus*, which are better resolved.

Parsimony analysis of the combined rDNA and mtDNA datasets from which *An. acanthotorynus* was excluded generated a single MPT with length = 5691, CI = 0.365 and RI = 0.534. Analysis using successive approximations character weighting identified a single, different tree with parsimony-informative length = 5693, CI = 0.365 and RI = 0.533 (Fig. 5). The single MPT is obviously more resolved than the strict consensus of the eight MPTs obtained from the analysis that included *An. acanthotorynus*, but it is nearly identical with one of those eight MPTs, differing only in the arrangement of basal taxa within *Cellia* and the position of *An. acanthotorynus*. Interestingly, however, in all other seven MPTs from the analysis that included *An. acanthotorynus*, *An. acanthotorynus* + *An. judithae*

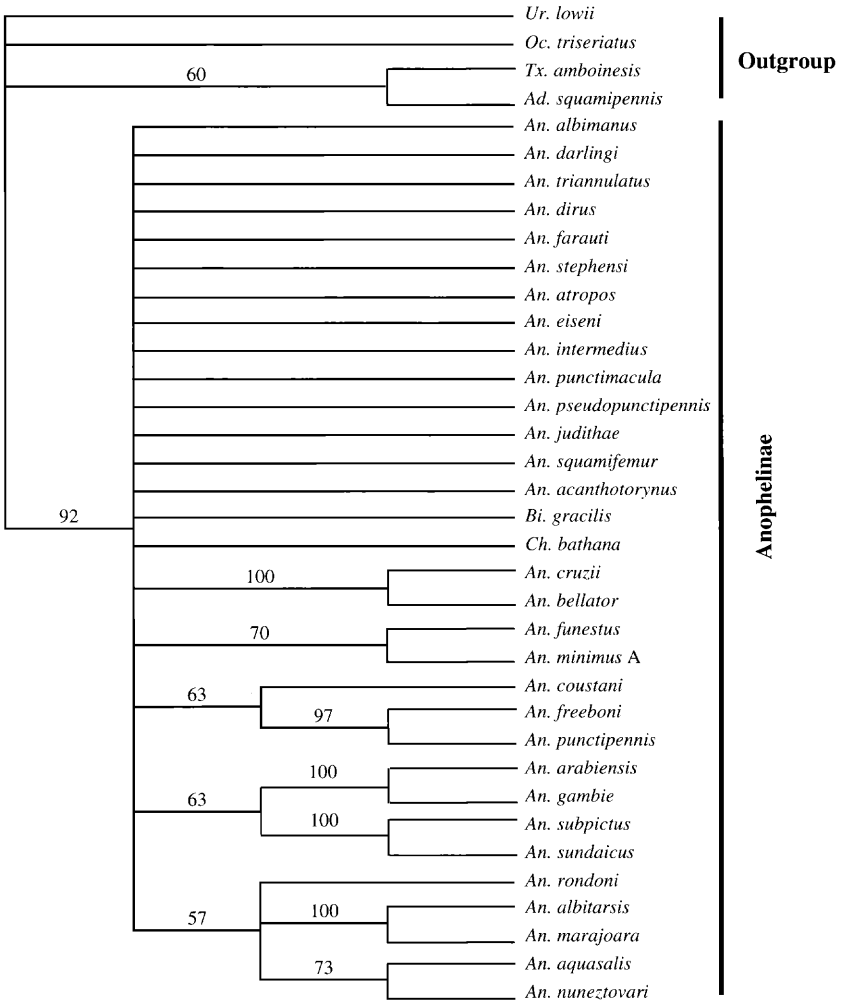

Fig. 3. The bootstrap consensus tree identified by unweighted parsimony analyses of the combined mtDNA (COI and COII) data. Numbers above branches indicate MP bootstrap proportions.

occupy basal positions in genus *Anopheles*. Analysis using successive approximations character weighting identified a single SWT (not one of the MPTs) that, except for the position of *An. acanthotorynus*, is identical to the SWT identified when *An. acanthotorynus* is included.

Likelihood. The SWT was evaluated under fifty-six models of evolution, and the most complex model available, GTR + I + Γ was found to be significantly better fitting than the next most complex alternative by both the likelihood ratio test ($P < 0.000001$) and the AIC. Likelihood analysis using GTR + I + Γ yielded a single most likely tree with a log likelihood of −29449.59348 (Fig. 6). Likelihood analysis with *An. acanthotorynus* excluded identified a single most likely tree with log likelihood of −28584.39475. This tree differs from the most likely tree found with *An. acanthotorynus* included only in the position of *An. acanthotorynus* and in the position of *An. triannulatus* within *Nyssorhynchus*.

Bootstrap analyses. Bootstrap support for most branches in the MP analysis increased when *An. acanthotorynus* was excluded; however, deeper relationships were weakly supported in both sets of MP analyses (< 50% bootstrap proportions). In contrast, the most basal relationships were better supported in ML bootstrap analyses. Due no doubt to the 'wildcard taxon' status of *An. acanthotorynus* on account of missing character data (Nixon & Wheeler, 1992), bootstrap support for nearly all branches is improved considerably when *An. acanthotorynus* is excluded versus when it is included in ML bootstrap analyses.

The single MP tree for the combined mtDNA and rDNA sequences with *An. acanthotorynus* excluded (Fig. 5) strongly supports the monophyly of Anophelinae as well as the grouping (*Anopheles* + *Bironella*) (Table 5). The topology of the MPT is largely identical to that of the tree identified by successive approximations character weighting except for the position of *Cellia*, which is the sister group of subgenus *Anopheles* in the MPT but is placed outside of the

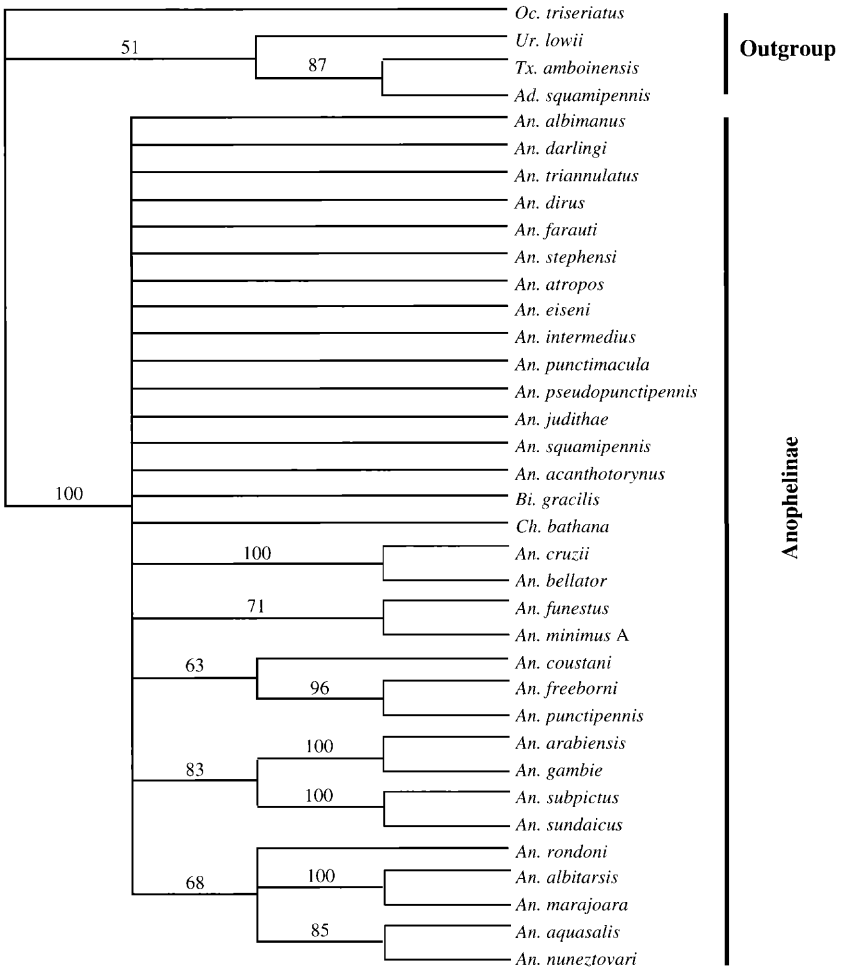

Fig. 4. The ML bootstrap consensus tree identified by maximum likelihood analyses of the combined mtDNA (COI and COII) data under the GTR + I + Γ model of nucleotide evolution. Numbers above branches indicate ML bootstrap proportions.

clade (*Anopheles* + ((*Bironella* + *Lophopodomyia*) + (*Kerteszia* + *Nyssorhynchus*))) in the latter tree. Bootstrap proportions for the monophyly of each of subgenera *Cellia*, *Kerteszia* and *Nyssorhynchus* are strong (all 100%), as is support for the sister-group relationship of *Kerteszia* + *Nyssorhynchus* (Table 5). The grouping *Lophopodomyia* + *Bironella* is well supported; however, the group formed by ((*Bironella* + *Lophopodomyia*) + (*Kerteszia* + *Nyssorhynchus*)) is only weakly supported. *Stethomyia* is recovered as the sister group of *An. judithae* within the clade formed by members of subgenus *Anopheles*; however, the grouping (*An. judithae* + *Stethomyia*) is weakly supported (Table 5). The sister-group relationship between *Lophopodomyia* and *Bironella* is not affected by the inclusion (96%) or exclusion (94%) of *Stethomyia*.

In ML analyses conducted with *Stethomyia* excluded (Fig. 6), support for a sister-group relationship between *Lophopodomyia* and *Bironella* is moderate, as is the support for the group ((*Bironella* + *Lophopodomyia*) + (*Kerteszia* + *Nyssorhynchus*)). In ML analyses in which *Stethomyia* is included, *Stethomyia* is reconstructed as the sister group of *Cellia* in the most likely tree, but is grouped with *Kerteszia* + *Nyssorhynchus* in the 50% majority rule bootstrap consensus tree; support for either relationship is low (Table 5). The sister-group relationship between *Lophopodomyia* and *Bironella* is moderately well supported; in contrast, the position of *Lophopodomyia* + *Bironella* as sister to *Nyssorhynchus* + *Kerteszia* is weakly supported (Table 5).

Evaluation of the relationship between Bironella *and genus* Anopheles

The controversial placement of *Bironella* within genus *Anopheles* was particularly scrutinized. ML bootstrap support for a position of *Bironella* within *Anopheles* is generally strong. In the combined analysis of the rDNA and mtDNA data (Fig. 6), for example, three nodes separate *Bironella* from the basal branch defining *Anopheles*, one of which is supported by a bootstrap proportion of 85% (Table 5). The S–H test (Shimodaira & Hasegawa, 1999) was used to compare the ML tree with suboptimal trees found under constrained searches. For this test, the nuclear ribosomal

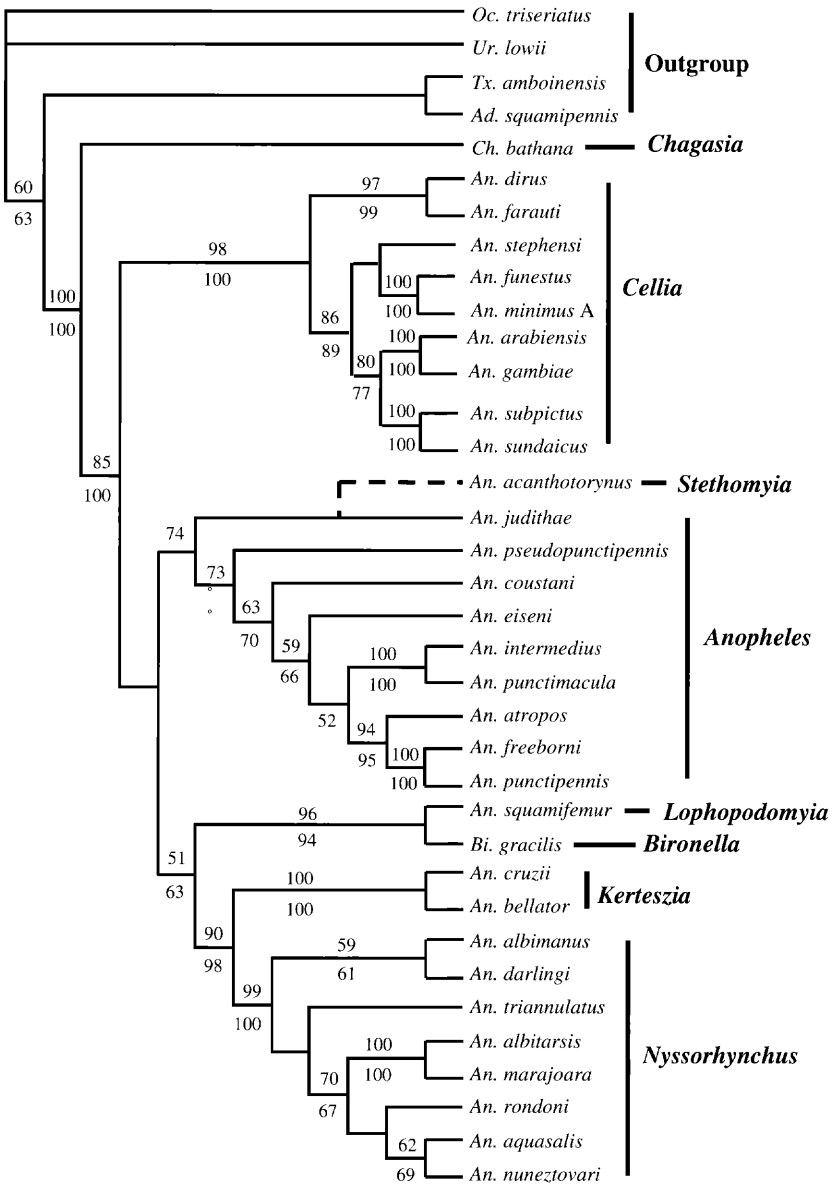

Fig. 5. The single tree identified by parsimony analyses using successive approximations weighting of the combined rDNA (18S and 28S) and mtDNA (COI and COII) data with *An.* (*Stethomyia*) *acanthotorynus* both included and excluded. Numbers above and below branches indicate MP bootstrap proportions obtained when *Stethomyia* was included and excluded from the analyses, respectively. Dashed line indicates the position of *Stethomyia*.

data were used, with *An. acanthotorynus* excluded, because this gave the clearest results as described above. To examine the support for placing *Bironella* within *Anopheles*, the ML tree found above was compared with a tree from an additional search conducted under the constraint that *Bironella* arises outside of *Anopheles*. The difference of 10.9582 log likelihood units was borderline significant ($P = 0.048$) by the S–H test, and so the possibility that *Bironella* is outside *Anopheles* is not strongly rejected by this test. Because the S–H test was inconclusive, the question of the position of *Bironella* was also examined using a likelihood ratio test for monophyly using parametric bootstrapping (Huelsenbeck *et al.*, 1996). This is a computationally intensive test that requires 200 separate ML analyses, so to shorten tree-search time some species were excluded because their sequences were near duplicates of those of their sister species. Specifically, *An. albitarsis*, *An. cruzii*, *An. arabiensis*, *An. subpictus*, *An. freeborni* and *An. intermedius* were excluded, leaving twenty-nine of the original thirty-six species. Additionally, all searches were constrained for well supported groups.

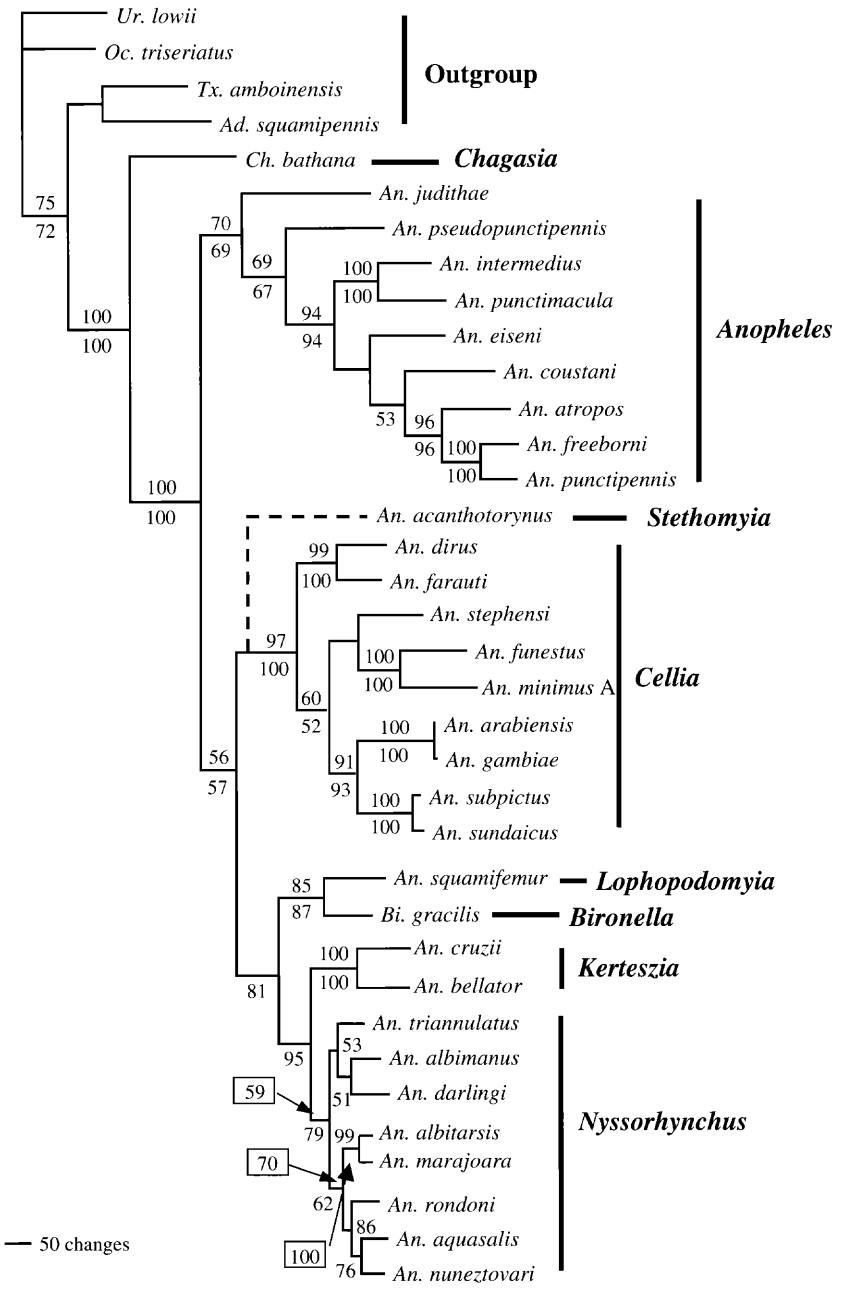

Fig. 6. The single tree identified by maximum likelihood analyses of the combined rDNA (18S and 28S) and mtDNA (COI and COII) data with *An.* (*Stethomyia*) *acanthotorynus* excluded under the GTR + I + Γ model of nucleotide evolution. This tree differs from the most likely tree found with *Stethomyia* included only in the position of *Stethomyia* and in the position of *An. triannulatus* within *Nyssorhynchus*. Numbers above and below branches indicate ML bootstrap proportions with *Stethomyia* included and excluded from the analyses, respectively. Dashed line indicates the position of *Stethomyia*.

The grouping (*Cellia* + (*Kerteszia* + *Nyssorhynchus*)) was constrained, as was the outgroup, including *Ch. bathana*. The positions of *Lophopodomyia* and the members of subgenus *Anopheles* were not otherwise constrained. The ML analyses for this test were conducted as described above for other ML analyses, except that empirically estimated base frequencies ('basefreq = empirical' in PAUP) were used.

The initial step in the likelihood ratio monophyly test was to conduct two ML analyses of the observed data, one in which the position of *Bironella* was not constrained, and one in which the position of *Bironella* was constrained to be outside of genus *Anopheles*. This constraint (or lack thereof) was imposed in addition to the constraints described above. The ML search in which the position of *Bironella* was not

constrained obtained, as before, a tree in which *Bironella* is the sister group of *Lophopodomyia*. The log likelihood difference between this tree and the tree found under the constraint that *Bironella* is not part of *Anopheles* is $\delta = 9.888$. Application of the S–H test indicates that there is no significant difference between these two trees ($P = 0.17$). However, the next step in the likelihood ratio monophyly test is to re-examine the significance of this difference using parametric bootstrapping. For this test, we consider the constrained tree, in which *Bironella* is outside of *Anopheles*, to be the null hypothesis, and we asked if it is reasonable to suppose that in the optimal (unconstrained) ML tree *Bironella* appears inside *Anopheles* only due to chance, i.e. as a result of the influence of random noise in the sequence data. To answer this question we need to know, on those occasions when the data produce such a result (i.e. when the ML tree indicates monophyly only due to chance), how much of a likelihood difference it is reasonable to expect, i.e. we need to obtain the expected distribution of the statistic for the null hypothesis of the likelihood difference. This null distribution was constructed by simulating data based on the null hypothesis: the constrained tree with its associated model ('parametric bootstrapping'). For each simulated dataset, a search was conducted for the best tree constrained to non-monophyly, and another search was conducted for the (unconstrained) ML tree. The statistic δ is the difference in log likelihood between them. Often the result of the two searches was the same tree, and so δ was zero. The null distribution of δ thus generated is shown in Fig. 7. Because the statistic $\delta = 9.888$ from the original data is well outside the null distribution, the null hypothesis can be rejected and it can be concluded that it is not plausible to suppose that *Bironella* is really outside of *Anopheles* and only appears to be derived within *Anopheles* on the optimal ML tree due to random noise. In conclusion, based on the null distribution obtained from parametric bootstrapping, there is significant phylogenetic signal that places *Bironella* as a derived group within genus *Anopheles*, and the monophyly of *Anopheles* exclusive of *Bironella* can be rejected. The probable explanation for the difference between the results of the non-parametric bootstrapping S–H test described above and the results of the likelihood-ratio test with parametric bootstrapping is that the latter test utilizes more information than the former, including information in the optimal model parameter values and what they imply about expected variance in the sequence data.

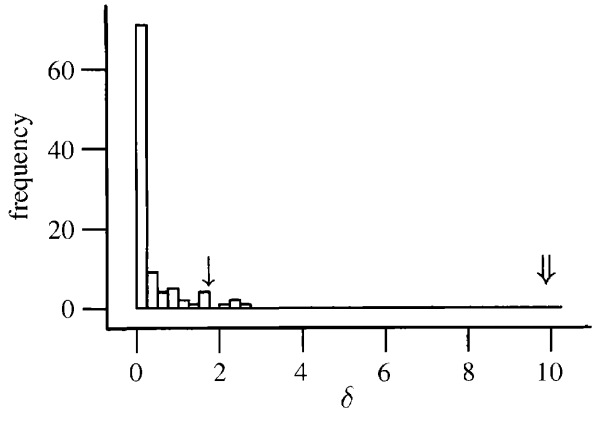

Fig. 7. Results of the likelihood ratio test of a derived position of *Bironella* within genus *Anopheles*. The single arrow indicates the position of the 95% confidence interval in the null distribution generated by parametric bootstrapping. Forty-eight of the parametric bootstrap replicates had $\delta = 0$. The double arrow indicates the position of the likelihood difference for the observed data. Because this value exceeds those found in all 100 parametric bootstrap replicates, a topology in which *Bironella* arises outside of *Anopheles*, i.e. in which *Bironella* retains separate generic status, is rejected.

The question of the phylogenetic position of *Bironella* can also be addressed using Bayesian analysis (Huelsenbeck & Ronquist, 2000; Lewis, 2001). In this analysis, a Markov Chain Monte Carlo (MCMC) method is used to explore parameter space (including tree space) in proportion to the posterior probability of the parameters and trees. Again, the nuclear ribosomal data were used, with *An. acanthotorynus* both excluded and included (Table 6). The posterior probability was approximately 1.0 for the splits which placed *Bironella* within *Anopheles*, which placed *Bironella* as the sister group of *Lophopodomyia* and which placed *Bironella* in a clade with *Nyssorhynchus* and *Kerteszia*. The posterior probability of topologies in which *Bironella* arises outside of *Anopheles* was estimated to be less than 0.001, although the posterior probability of topologies in which a clade containing both *Bironella* and *An. squamifemur* arises outside of the rest of *Anopheles* was estimated to be about 0.05.

Table 6. Clade support for ribosomal and combined ribosomal plus mitochondrial sequences using Bayesian analyses. Numbers are posterior probabilities.

	18S + 28S	rDNA + mtDNA
Sister group		
Chagasia and *Anopheles*	NA*/NA	NA/NA
Monophyly		
Anopheles (including *Bironella*)	1.0*/1.0	1.0*/1.0
Subgenera		
Anopheles	1.0*/1.0	0.98*/0.97
Kerteszia	1.0*/1.0	1.0*/1.0
Cellia	1.0*/1.0	1.0*/1.0
Nyssorhynchus	0.04*/0.09	1.0*/1.0
Other groupings		
(*Lophopodomyia*, *Bironella*)	0.97*/1.0	1.0*/1.0
(*Stethomyia*, *Cellia*)	0.82*/NA	1.0*/NA
(*Kerteszia*, *Nyssorhynchus*)	0.99*/1.0	1.0*/1.0
BLNK	0.83*/0.94	1.0*/1.0
Cellia, BLNK	0.01*/0.17	0.89*/0.69
Anopheles, *Cellia*	0.72*/0.73	0.00*/C

C, contradicted by an alternative relationship appearing on bootstrap tree. NA = not applicable. *, *Stethomyia* included/*Stethomyia* excluded. BLNK = *Bironella*, *Lophopodomyia*, *Nyssorhynchus*, *Kerteszia*.

Evaluation of the relationship between Nyssorhynchus *and* Kerteszia

In the ML tree based on ribosomal sequences, *Kerteszia* arises from within *Nyssorhynchus*, as sister to *An. triannulatus* (Fig. 2). Support for this result was scrutinized. Whereas ML support using ribosomal data for the clade *Kerteszia* + *Nyssorhynchus* is moderate (81% bootstrap proportion), support for *Kerteszia* as sister to *An. triannulatus* is weak (52%). When the ML search was repeated under the constraint that *Nyssorhynchus* is monophyletic, a tree was found that is only 1.7311 log likelihood units worse than the optimal (unconstrained) ML tree, a difference that is not significant ($P = 0.225$) based on the S–H test. Bayesian analysis using these data strongly supported monophyly of the group *Kerteszia* + *Nyssorhynchus* (Table 6), with a posterior probability of 1.0. However, little support was found for any particular split associating *Kerteszia* with any particular member of *Nyssorhynchus* (posterior probability ≤ 0.32). The posterior probability of monophyletic *Nyssorhynchus* was only 0.09. This analysis corroborates the bootstrap analyses and the S–H test, all of which show that the ribosomal data are unable to reliably resolve the relationships among species of *Kerteszia* and *Nyssorhynchus*.

Whereas the ribosomal data alone failed to reliably resolve subgenera in this part of the tree (Figs 1, 2), combined mitochondrial and ribosomal data indicate a monophyletic *Nyssorhynchus* (Figs 5, 6). ML support is fair (79%) when *Stethomyia* is excluded, and using these same data, a Bayesian analysis calculates the posterior probability for the monophyly of *Nyssorhynchus* to be 0.999. In both the ML and Bayesian analyses, *Kerteszia* is the sister of *Nyssorhynchus*. Additionally, using the combined mtDNA and rDNA data, a trend in the resolution of members of *Nyssorhynchus* was found, with *An. triannulatus*, *An. darlingi* and *An. albimanus* associated (in an adjacent pectinate if not monophyletic relationship) and the other five taxa monophyletic (Figs 5, 6).

Basal resolution of the subgenera of genus Anopheles

In the ML tree based on ribosomal data, the relationship of the subgenera is ((*Cellia* + *Anopheles*) + (BLNK)) (Fig. 2), where BLNK refers to a clade consisting of *Bironella* + *Lophopodomyia* + *Nyssorhynchus* + *Kerteszia*. As this grouping is not consistently obtained in our analyses, support for alternative basal relationships of subgenera was examined. For example, *Cellia* was placed outside the clade formed by (*Anopheles* + (BLNK)) in the parsimony trees resulting from successive approximations character weighting analyses of both ribosomal (Fig. 1), and combined rDNA plus mtDNA datasets (Fig. 5). In contrast, *Cellia* exchanged positions with subgenus *Anopheles* in the topology resulting from ML analyses of the combined rDNA plus mtDNA data, in which *Anopheles* arises outside of the clade consisting of (*Cellia* + (BLNK)) (Fig. 6).

Table 7. Shimodaira–Hasegawa RELL test of basal arrangements of the subgenera of *Anopheles*[1].

Tree	LnL	Δ	*P*
Cellia, *Anopheles*	−12779. 1934	0.0000	(maxLike)
Anopheles, BLNK[2]	−12781.4773	2.2839	0.389
Cellia, BLNK	−12780.9916	1.7982	0.444

[1]There were 1000 bootstrap replicates. The *P* values are based on the position of Δ in the bootstrap replicates.
[2]BLNK = *Bironella*, *Lophopodomyia*, *Nyssorhynchus*, *Kerteszia*.

We examined this question of basal subgeneric relationships with the S–H test. We searched for the best trees that constrained the basal relationships as ((*Cellia* + BLNK) + *Anopheles*) or ((*Anopheles* + BLNK) + *Cellia*), and examined the differences in log likelihoods between these suboptimal trees and the ML tree ((*Cellia* + *Anopheles*) + BLNK). The results show that the other two constrained arrangements are not significantly worse-fitting to the data than is the optimal (unconstrained) ML tree (Table 7).

The consensus tree from a Bayesian analysis also supports the arrangement ((*Cellia* + *Anopheles*) + BLNK), consistent with the ML tree. The posterior probability was 0.73 for *Anopheles* + *Cellia*, 0.17 for *Cellia* + BLNK, and 0.088 for *Anopheles* + BLNK (Table 6). These probabilities reflect the same order of preference as found in the ML analyses but are more definitive.

Differences in results of analyses of 18S and 28S vs rDNA and mtDNA datasets

A comparison of the bootstrap values indicates that the major disagreement between the two datasets, 18S and 28S versus the combined mtDNA and rDNA, is restricted to the relationships between the three major clades: subgenus *Anopheles*, subgenus *Cellia* and the clade consisting of *Lophopodomyia* + *Bironella* + *Kerteszia* + *Nyssorhynchus*. The ML analyses of the 18S and 28S dataset suggest that *Cellia* and *Anopheles* are sister groups (Fig. 2), whereas ML analyses of the combined rDNA and mtDNA dataset suggests that (*Cellia* + (*Lophopodomyia* + *Bironella* + *Kerteszia* + *Nyssorhynchus*)) are sister groups (Fig. 5). Bootstrap support for either set of relationships is negligible, however, at < 50% for the former and 57% for the latter. For the data presented here, this problem of basal relationships within Anophelinae is best resolved by the Bayesian analyses described above, which support the grouping ((*Lophopodomyia* + *Bironella* + *Kerteszia* + *Nyssorhynchus*) and (*Cellia* + *Anopheles*)).

The effect of including/excluding An. acanthotorynus

Except for the presence of *An. acanthotorynus*, identical 50% majority rule consensus trees are obtained from bootstrap

analyses of the combined rDNA data, including and excluding *An. acanthotorynus*. A comparison of bootstrap proportions on the two trees (Fig. 2) suggests that the inclusion of *An. acanthotorynus* decreases confidence in the most basal branches within genus *Anopheles*, due to the ambiguous position of *An. acanthotorynus* caused by missing 28S sequence data. Specifically, bootstrap support for the branch uniting *Lophopodomyia*, *Bironella*, *Kerteszia* and *Nyssorhynchus* increases from < 50% to 69% when *An. acanthotorynus* is excluded; likewise, support for the branch uniting *Lophopodomyia* and *Bironella* increases from 75% to 92% when *An. acanthotorynus* is excluded.

We note, however, that bootstrap support for the position of *An. acanthotorynus* as the sister group of *Cellia* is quite high in the MP analysis of the combined 18S and 28S data (94%). This relationship is largely supported by the indel coding, and is thus not recovered in the ML analyses, which are unable to make use of the indel information.

Discussion

The monophyly of Anophelinae is unequivocally supported by phylogenetic analysis of morphological characters (Harbach & Kitching, 1998; Sallum *et al.*, 2000), previous molecular data (Foley *et al.*, 1998; Krzywinski *et al.*, 2001a,b) and analyses of nearly all the partitions of the present molecular data (Table 5). The only exception is the maximum likelihood analysis of the 18S data, which only weakly supports this relationship.

The results of the present molecular analyses agree with the current classification of Anophelinae, giving support for the non-monophyly of genus *Anopheles*. The placement of *Bironella* within genus *Anopheles* contradicts the results of other analyses (Besansky & Fahey, 1997; Foley *et al.*, 1998; Harbach & Kitching, 1998), including the results of Krzywinski *et al.* (2001b) for the nuclear white gene. Support for the monophyly of genus *Anopheles* (including *Bironella*) is strongly supported under both parsimony and likelihood by all rDNA analyses (18S, 28S and 18S plus 28S), and by analyses of the combined rDNA and mtDNA data. Although analyses of the mtDNA data in isolation provide only weak support for this conclusion, it appears in the optimal tree for the ML analyses and in the optimal trees for the MP amino acid analyses, but not in the optimal tree for the unweighted MP analyses of mtDNA nucleotides. Foley *et al.* (1998) found strong support for the monophyly of the genera *Bironella* and *Anopheles* using successive approximations character weighting analyses. However, this result may be an artefact of taxon sampling (Hillis, 1998), as the scope of that study was to estimate phylogenetic relationships among Australasian species of *Anopheles*, and thus no representative of Neotropical subgenera was included in the analyses. The significance of the placement of *Bironella* within genus *Anopheles* was examined using constraint analyses and statistical tests. In particular, the likelihood ratio test for monophyly using parametric bootstrapping (Huelsenbeck *et al.*, 1996) for the rDNA data showed that the monophyly of genus *Anopheles* relative to *Bironella* can be rejected. Likewise, the results of the analyses of rDNA and combined rDNA plus mtDNA data using Bayesian analysis also support this conclusion. This hypothesis of non-monophyly of genus *Anopheles* is also supported by analyses of sequence data from the ND5 gene alone, combined ND5 plus D2, G_6pd alone, and combined G_6pd, ND5, D2 and white genes, all of which placed *Bironella* within *Anopheles* (Krzywinski *et al.*, 2001a,b) with varying degrees of support. This result of a nested position of *Bironella* within *Anopheles* cannot be explained simply as an artefact of taxon or character sampling because it is supported by data from diverse sources, including molecular data from various gene fragments and morphological data (Sallum *et al.*, 2000; Krzywinski *et al.*, 2001a,b)

The results of the present analyses, which place *Bironella* as the sister group to *Lophopodomyia* in a clade that also contains *Nyssorhynchus* and *Kerteszia*, contradict the previously hypothesized position of *Bironella* within a clade formed by representatives of the subgenera *Anopheles*, *Stethomyia*, *Lophopodomyia* and *Bironella*, suggested by Sallum *et al.* (2000). In fact, Sallum *et al.* (2000) regarded *Bironella* as an informal group within genus *Anopheles* because its position within the larger clade was unresolved and unsupported by MP bootstrap proportions above 50%. Based on the totality of their results, a close phylogenetic relationship between *Bironella* and subgenus *Anopheles* was rejected by Krzywinski *et al.* (2001a,b). Judging the combined evidence produced by all molecular studies carried out to date, and given the demonstrated morphological similarity between *Bironella* and *Anopheles* (Sallum *et al.*, 2000), the weight of evidence favours the current hypothesis of the paraphyly of genus *Anopheles* with respect to *Bironella*. Moreover, the results of statistical analyses of the rDNA data (S–H tests, log likelihood ratio monophyly tests and Bayesian analyses) all indicate a reconstruction in which the common ancestor of the group consisting of *Bironella*, *Lophopodomyia*, *Nyssorhynchus* and *Kerteszia* diverged earliest, and in which the separate ancestors of *Cellia* and *Anopheles* diverged later. Given the disagreement among various molecular data, the exact position of *Bironella* within genus *Anopheles* remains unresolved. However, the results of this study suggest that *Bironella* may be plausibly regarded as a subgenus of *Anopheles*. As this study included only one species of *Bironella* and few species of subgenus *Anopheles*, this hypothesis of a subgeneric status for *Bironella* requires further testing using broader taxon sampling.

Subgenus Anopheles

The monophyly of subgenus *Anopheles* has remained in doubt because the results of the morphology-based phylogeny of Sallum *et al.* (2000) disagree with the hypothesis of monophyly suggested by Krzywinski *et al.* (2001a,b). In the current study, monophyly of subgenus *Anopheles* is

only moderately well supported. Support for the monophyly of subgenus *Anopheles* is almost entirely due to phylogenetic signal residing in the 28S dataset but, as can be seen from examining the bootstrap proportions in Table 5, some support for this group also resides in the 18S data. It is important to note that, except for *An. coustani*, all other members of subgenus *Anopheles* included in this study are from the Neartic and Neotropical Regions, and thus the apparent monophyly of subgenus *Anopheles* may be an artefact of taxon sampling that will break down when more taxa are added, especially those from the Old World (specifically, Africa, Australasia and Southeast Asia). Similarly, Krzywinski *et al.* (2001a,b) included subgenus *Anopheles* species, all of which were Neartic/Neotropical except for *An. coustani*. In contrast, Foley *et al.* (1998) proposed that subgenera *Anopheles* and *Cellia* are paraphyletic, because *An. quadrimaculatus* is nested within *Cellia*. Based on analyses of mtDNA in the present study, these conclusions may be an artefact of weak phylogenetic signal in COII for this level of analysis. The Arribalzagia Series of subgenus *Anopheles* appears to be a monophyletic group (Wilkerson & Peyton, 1990; Sallum *et al.*, 2000), which accords with all the results of the present study. The Anopheles Series was previously found to be paraphyletic (Sallum *et al.*, 2000), a result that also accords with all the results of the present study, as its species are dispersed throughout the clade. The placement of *An. coustani*, of the Myzorhynchus Series, is not consistent in all the analyses. Relationships among the Myzorhynchus, Arribalzagia and Anopheles Series remain unresolved because the present data do not provide unambiguous support for any relationships. Moreover, the exact placement of subgenus *Anopheles* within genus *Anopheles* remains unresolved as a result of poor resolution of basal relationships within trees.

Subgenus Cellia

The results of the present study, with the exception of the mtDNA analyses, strongly support subgenus *Cellia* as a distinct monophyletic lineage within genus *Anopheles*. In contrast, Foley *et al.* (1998) proposed the paraphyly of *Cellia* relative to subgenus *Anopheles*. Because the monophyly of *Cellia* is supported both by morphological characters (Sallum *et al.*, 2000) and by various gene sequences (Krzywinski *et al.*, 2001a,b), it is reasonable to conclude that *Cellia* is a monophyletic group within genus *Anopheles*. *Cellia* has been traditionally divided into six series, Cellia, Neocellia, Myzomyia, Pyretophorus, Paramyzomyia and Neomyzomyia. Except for the Cellia Series, all the remaining series were previously found to be paraphyletic or polyphyletic based on morphological characters (Sallum *et al.*, 2000). In contrast, all of the analyses of the rDNA and combined rDNA plus mtDNA data reported here support the monophyly of the Pyretophorus, Myzomyia, Neocellia and Neomyzomyia Series. The monophyly of the Pyretophorus Series is consistent with the hypothesis of Anthony *et al.* (1999). Foley *et al.* (1998) also reported support for the monophyly of the Pyretophorus, Neomyzomyia and Myzomyia Series, but not for Neocellia, which was regarded as paraphyletic. As with the position of subgenus *Anopheles*, the available data do not confidently clarify the position of the subgenus *Cellia* within genus *Anopheles*.

Subgenus Stethomyia

Because nucleotide data from the 18S rDNA and mtDNA gene regions are apparently insufficient to strongly link *Stethomyia* (represented by only one species, *An. acanthotorynus*, in the analyses) to any other species or group of species, *Stethomyia* behaves as a 'wildcard taxon' (Nixon & Wheeler, 1992) in bootstrap analyses, inserting at multiple positions in the topology in near-optimal trees and thereby eroding bootstrap support for numerous branches. To gauge this effect, analyses were conducted with *Stethomyia* both included and excluded. The strongest evidence for the position of *Stethomyia* comes from parsimony analyses of the 18S rDNA data, in which *Stethomyia* groups with *Cellia* with a 90% bootstrap proportion; in ML analyses, support for this grouping is $<50\%$. This disparity is explained by the fact that, rather than from nucleotide sequences, most of the support for the *Stethomyia* and *Cellia* grouping derives from shared multiple-position indels, coded using the simple indel coding method of Simmons & Ochoterena (2000); these indel characters were necessarily ignored in the likelihood analyses. When the 18S data are combined with the 28S rDNA data and analysed with parsimony, support for the grouping *Stethomyia* + *Cellia* increases to 94% in spite of the fact that *Stethomyia* entirely lacks 28S sequence data; again, this support drops to $<50\%$ when the combined nucleotide data (but not the indel data) are analysed using likelihood. Although bootstrap support for nearly all relationships is low ($<50\%$), it is worth noting that *Stethomyia* again groups with *Cellia* (arising within that subgenus) in the optimal trees resulting from both MP and ML analyses of the mtDNA data. In spite of this, and for reasons we do not fully understand, this relationship does not appear in any of the trees produced by the combined rDNA and mtDNA analyses. In fact, the tree resulting from the successive-approximations character-weighted parsimony analysis of the combined data places *Stethomyia* as the sister group of *An. judithae*, with the combined taxa forming the sister group to the remainder of subgenus *Anopheles*. These results suggest that the alternative grouping of *Stethomyia* with either *Cellia* or *An. judithae* in different MP analyses may be due to a combination of long-branch attraction between the most derived taxa included in the analyses as well as to conflicting phylogenetic signal within the rDNA data.

Such a sister-group relationship between *Stethomyia* and *Cellia* has not been suggested previously, and it is strongly supported only by our rDNA parsimony analyses. Certainly there is no morphological evidence to support

this hypothesis. Sallum *et al.* (2000) showed *Stethomyia* to be the sister group of *Bironella*, with both clades arising well within subgenus *Anopheles*. Based on the incongruence between the morphological and molecular results, elevation of *Stethomyia* to the subgeneric level would certainly be premature, because its position continues to remain unclear.

Subgenus Lophopodomyia

Although a relationship between the Neotropical subgenus *Lophopodomyia* and the Australasian genus *Bironella* has never been suggested, the results of the present analyses weakly support the association of these taxa within a larger clade that also includes *Nyssorhynchus* and *Kerteszia*. Because 28S rDNA data could not be obtained for *Bironella* and because the mtDNA data contain little phylogenetic information at the level of this study, it is clear that most of the character support for the sister-group relationship of *Lophopodomyia* + *Bironella* resides in the 18S rDNA data. In spite of the fact that no 28S data exist for *Bironella*, the interaction of the 18S and 28S datasets increases support for this grouping in combined ML analyses. In ML analyses of the rDNA and the combined rDNA and mtDNA data, *Lophopodomyia* is consistently placed as the sister group to the clade *Nyssorhynchus* + *Kerteszia* with consistently high bootstrap support (92% and 90%, respectively). Subgenus *Lophopodomyia* is similarly found to be the sister group of *Nyssorhynchus* + *Kerteszia* in likelihood analyses of both the nuclear 'white' gene and of combined DNA data from multiple genes (Krzywinski *et al.*, 2001b). Thus, *Lophopodomyia* appears to be a monophyletic group separate from subgenus *Anopheles*.

Subgenera Kerteszia *and* Nyssorhynchus

Monophyly of *Nyssorhynchus* is a consistent feature of trees produced by MP analyses, and is supported by high bootstrap proportions (Table 5). Monophyly of *Nyssorhynchus* is also indicated by ML analyses using combined rDNA and mtDNA data. ML analyses with rDNA alone do not reliably confirm or reject the monophyly of *Nyssorhynchus*. The optimal trees resulting from ML analyses of the combined rDNA and mtDNA data with *Stethomyia* both included and excluded reconstruct *Kerteszia* and *Nyssorhynchus* as sister taxa. This result is largely congruent with the results of a cladistic analysis of morphological characters (Sallum *et al.*, 2000) and with those of previous molecular analyses (Krzywinski *et al.*, 2001a,b). Within *Nyssorhynchus*, phylogenetic relationships remain unclear. However, the paraphyly of the Argyritarsis and Albimanus Sections are confirmed. It is important to note that the problem of relationships of species within *Nyssorhynchus* is one of the few areas for which the mtDNA sequence data contribute decisive information, suggesting, as do the low genetic distances between the DNA sequences of these species, that these species are the result of relatively recent divergence.

Acknowledgements

We thank Bruce Harrison, Mark Benedict, Nora Besansky, Pollie Rueda, Sylvie Manguin, Thomas Gaffigan and Thomas Zavortink for providing or helping to collect specimens. We extend our thanks to C. Huddleston for his valuable assistance in collecting sequences, and the anonymous reviewers who helped us to improve this paper with their constructive comments. This research was partially supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Brazil (MAMS), grant no. 97/8792-9; the Department of Entomology of the Walter Reed Army Institute of Research; and the Laboratory of Molecular Systematic of the Smithsonian Institution. During the course of this research, T.R.S. was supported by National Science Foundation Grant DEB 9707209 and Smithsonian Scholarly Studies Grant 140202-3340-410.

References

Abouheif, E., Zardoya, R. & Meyer, A. (1998) Limitations of metazoan 18S rRNA sequence data: implications for reconstructing a phylogeny of the animal kingdom and inferring the reality of the Cambrian explosion. *Journal of Molecular Evolution*, **47**, 394–405.

Anthony, T.G., Harbach, R.E. & Kitching, I.J. (1999) Phylogeny of the Pyretophorus Series of *Anopheles* subgenus *Cellia* (Diptera: Culicidae). *Systematic Entomology*, **24**, 193–205.

Belshaw, R., Fitton, M., Hernion, E., Gimeno, C. & Quicke, D.L.J. (1998) A phylogenetic reconstruction of the Ichneumonoidea (Hymenoptera) based on the D2 variable region of 28S ribosomal RNA. *Systematic Entomology*, **23**, 109–123.

Besansky, N.J. & Fahey, G.T. (1997) Utility of the white gene in estimating phylogenetic relationships among mosquitoes (Diptera: Culicidae). *Molecular Biology and Evolution*, **14**, 442–454.

Bull, J.J., Huelsenbeck, J.P., Cunningham, C.W., Swofford, D.L. & Wadell, P.J. (1993) Partitioning and combining data in phylogenetic analysis. *Systematic Biology*, **42**, 384–397.

Carpenter, J.M. (1988) Choosing among equally parsimonious cladograms. *Cladistics*, **4**, 291–296.

Chippindale, P.T., Dave, V.K., Whitmore, D.H. & Robinson, J.V. (1999) Phylogenetic relationships of North American damselflies of the genus *Ischnura* (Odonata: Zygoptera: Coenagrionidae) based on sequences of three mitochondrial genes. *Molecular Phylogenetics and Evolution*, **11**, 110–121.

Crozier, R.H. & Crozier, Y.C. (1993) The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization. *Genetics*, **133**, 97–117.

Cunningham, C.W. (1997) Can three incongruence tests predict when data should be combined? *Molecular Biology and Evolution*, **14**, 733–740.

De Peer, Y., Jansen, J., De Rijk, P. & De Wachter, R. (1997) Database on the structure of small ribosomal subunit RNA. *Nucleic Acids Research*, **25**, 111–116.

DeSalle, R. & Brower, A.V.Z. (1997) Process partitions, congruence and the independence of characters: inferring

relationships among closely related Hawaiian *Drosophila* from multiple gene regions. *Systematic Biology*, **46**, 751–764.

Eernisse, D.J. & Kluge, A.G. (1993) Taxonomic congruence versus total evidence, and amniote phylogeny inferred from fossils, molecules, and morphology. *Molecular Biology and Evolution*, **10**, 1170–1195.

Farris, J.S. (1969) A successive approximations approach to character weighting. *Systematic Zoology*, **18**, 374–385.

Farris, J.S., Kallersjo, M., Kluge, A.G. & Bult, C. (1995) Constructing a significance test for incongruence. *Systematic Biology*, **44**, 570–572.

Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, **39**, 783–791.

Felsenstein, J. (1996) Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. *Methods in Enzymology*, **266**, 418–427.

Foley, D.H., Bryan, J.H., Yeats, D. & Saul, A. (1998) Evolution and systematics of *Anopheles*: insights from a molecular phylogeny of Australasian mosquitoes. *Molecular Phylogenetics and Evolution*, **9**, 262–275.

Frati, F., Simon, C., Sullivan, J. & Swofford, D.L. (1997) Evolution of the mitocondrial cytochrome oxidase II gene in Collembola. *Journal of Molecular Evolution*, **44**, 145–158.

Gimeno, C., Belshaw, R. & Huike, D.L.J. (1997) Phylogenetic relationships of the Alysinae/Opiinae (Hymenoptera: Branconidae) and the utility of cytochrome b, 16S and 28S, D2 rRNA. *Insect Molecular Biology*, **6**, 273–284.

Giribet, G. & Ribera, C. (1998) The position of arthropods in animal kingdom: a search for a reliable outgroup for internal arthropod phylogeny. *Molecular Phylogenetics and Evolution*, **9**, 481–488.

Giribet, G. & Ribera, C. (2000) A review of arthropod phylogeny: new data based on ribosomal sequences and direct character optimization. *Cladistics*, **16**, 204–231.

Goldman, N., Anderson, J.P. & Rodrigo, A.G. (2000) Likelihood-based test of topologies in phylogenetics. *Systematic Biology*, **49**, 652–670.

Harbach, R.E. (1994) Review of the internal classification of the genus *Anopheles* (Diptera: Culicidae): the foundation for comparative systematics and phylogenetic research. *Bulletin of Entomological Research*, **84**, 331–342.

Harbach, R.E. & Kitching, I.J. (1998) Phylogeny and classification of the Culicidae (Diptera). *Systematic Entomology*, **23**, 327–370.

Hillis, D.M. (1998) Taxonomic sampling, phylogenetic accuracy, and investigator bias. *Systematic Biology*, **47**, 3–8.

Huelsenbeck, J.P. & Ronquist, F. (2000) *Mr Bayes: Bayesian inference of phylogeny*. Distributed by the Author. Department of Biology, University of Rochester, Rochester.

Huelsenbeck, J.P., Hillis, D.M. & Nielsen, R. (1996) A likelihood-ratio test of monophyly. *Systematic Biology*, **45**, 546–558.

Klompen, J.S.H., Black, I.V.W.C., Keirans, J.E. & Norris, D.E. (2000) Systematic and biogeography of hard ticks, a total evidence approach. *Cladistics*, **16**, 79–102.

Knight, K.L. & Stone, A. (1977) *A Catalog of the Mosquitoes of the World (Diptera: Culicidae). The Thomas Say Foundation*. 2nd edn. Entomological Society of America, Baltimore.

Krzywinski, J., Wilkerson, R.C. & Besansky, N.J. (2001a) Evolution of mitochondrial and ribosomal gene sequences in Anophelinae (Diptera: Culicidae): implications for phylogeny reconstrution. *Molecular Phylogenetics and Evolution*, **18**, 479–487.

Krzywinski, J., Wilkerson, R.C. & Besansky, N.J. (2001b) Toward understanding Anophelinae (Diptera: Culicidae) phylogeny: insights from nuclear single copy genes and the weight of evidence. *Systematic Biology*, **50**, 540–556.

Lewis, P.O. (2001) Phylogenetic systematics turns over a new leaf. *Trends in Ecology and Evolution*, **16**, 30–37.

Lunt, D.H., Zhang, D.X., Szymura, J.M. & Hewitt, G.M. (1996) The insect cytochrome oxidase I gene: evolutionary patters and conserved primers for phylogenetic studies. *Insect Molecular Biology*, **5**, 153–165.

Maddison, D.R. (1991) The discovery and importance of multiple islands of most-parsimonious trees. *Systematic Zoology*, **40**, 315–328.

Maddison, D.R., Baker, M.D. & Ober, K.A. (1999) Phylogeny of carabid beetles as infered from 18S ribosomal DNA (Coleoptera: Carabidae). *Systematic Entomology*, **24**, 103–138.

Maddison, D.R. & Maddison, W.P. (2000) *MacClade 4. Analysis of Phylogeny and Character Evolution*. Sinauer Associates, Sunderland, Massachusetts.

Miller, B.R., Crabtree, M.B. & Savage, H.M. (1997) Phylogenetic relationships of the Culicomorpha inferred from 18S and 5.8S ribosomal DNA sequences (Diptera: Nematocera). *Insect Molecular Biology*, **6**, 105–114.

Mitchell, A., Mitter, C. & Regier, J.C. (2000) More taxa or more characters revisited: combining data from nuclear protein-encoding genes for phylogenetic analyses of Noctuoidea (Insecta: Lepidoptera). *Systematic Biology*, **49**, 202–224.

Nixon, K.C. & Carpenter, J.M. (1996) On simultaneous analysis. *Cladistics*, **12**, 221–241.

Nixon, K.C. & Wheeler, Q.D. (1992) Extinction and the origin of species. *Extinction and Phylogeny* (ed. by M. J. Novacek and Q. D. Wheeler), pp. 119–143. Columbia University Press, New York.

Palumbi, S.R. (1996) Nucleic acids II: the polymerase chain reaction. *Molecular Systematics* (ed. by D. M. Hillis, C. Moritz and B. K. Mable), pp. 205–247. 2nd edn. Sinauer Associates, Sunderland, Massachusetts.

Pawlowski, J., Szadziewski, R., Kmieciak, D., Fahrni, J. & Bittar, G. (1996) Phylogeny of the infraorder Culicomorpha (Diptera: Nematocera) based on 28S RNA gene sequences. *Systematic Entomology*, **21**, 167–178.

Porter, C.H. & Collins, F.H. (1996) Phylogeny of Neartic members of the *Anopheles maculipennis* species group derived from the D2 variable region of 28S ribosomal RNA. *Molecular Phylogenetics and Evolution*, **6**, 178–188.

Posada, D. & Crandall, K.A. (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics*, **14**, 817–818.

de Queiroz, A., Donoghue, M.J. & Kim, J. (1995) Separate versus combined analysis of phylogenetic evidence. *Annual Review of Ecology and Systematic*, **26**, 67–681.

Rodríguez, F., Oliver, J.L., Marin, A. & Madina, J.R. (1990) The general stochastic model of nucleotide substitution. *Journal of Theoretical Biology*, **142**, 485–501.

Sallum, M.A.M., Schultz, T.R. & Wilkerson, R.C. (2000) Phylogeny of Anophelinae based on morphological characters. *Annals of the Entomological Society of America*, **93**, 745–775.

Shimodaira, H. & Hasegawa, M. (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution*, **16**, 1114–1116.

Simmons, M.P. & Ochoterena, H. (2000) Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology*, **49**, 369–381.

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H. & Flook, P. (1994) Evolution, weighting and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America*, **87**, 651–701.

Swofford, D.L. (1998) *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods), Version* 4. Sinauer Associates, Sunderland, Massachusetts.

Swofford, D.L., Olsen, G.J., Wadell, P.J. & Hillis, D.M. (1996) Phylogenetic inference. *Molecular Systematics* (ed. by D. M. Hillis, C. Moritz and B. K. Mable), pp. 407–514. 2nd edn. Sinauer Associates, Sunderland, Massachusetts.

Tamura, K. & Nei, M. (1993) Estimation of the number of nucleotide substitution in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, **10**, 512–526.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. & Higgins, D.G. (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, **24**, 4876–4882.

Thompson, J.D., Higgins, D.G., Gibson, T.J. (1996) ClustalW, 1.6. http://biomaster.uio.no/clustalw.html.

Thornton, J.W. & DeSalle, R. (2000) A new method to localize and test the significance of incongruence: detecting domain shuffling in the nuclear receptor superfamily. *Systematic Biology*, **49**, 183–201.

Wetterer, J.K., Schultz, T.R. & Meier, R. (1998) Phylogeny of fungus-growing ants (tribe Attini) based on mtDNA sequence and morphology. *Molecular Phylogenetics and Evolution*, **9**, 42–47.

Wiens, J.J. (1998) Does adding characters with missing data increase or decrease phylogenetic accuracy? *Systematic Biology*, **47**, 625–640.

Wiens, J.J. & Reeder, T.W. (1995) Combining data sets with different numbers of taxa for phylogenetic analysis. *Systematic Biology*, **44**, 548–558.

Wilkerson, R.C., Parsons, T.J., Albright, D.G., Klein, T.A. & Braun, M.J. (1993) Random amplified polymorphic DNA (RAPD) markers readily distinguish cryptic mosquito species (Diptera: Culicidae: *Anopheles*). *Insect Molecular Biology*, **1**, 205–211.

Wilkerson, R.C. & Peyton, E.L. (1990) Standardized nomenclature for costal wing spots of the genus *Anopheles* and other spotted-wing mosquitoes (Diptera: Culicidae). *Journal of Medical Entomology*, **27**, 208–224.

Winnepenninckx, B.M.H., Reid, D.G. & Backeljau, T. (1998) Performance of 18S rRNA in littorinid phylogeny (Gastropoda: Caenogastropoda). *Journal of Molecular Evolution*, **47**, 586–596.

Yoder, A.D., Irwin, J.A. & Payseur, B.A. (2001) Failure of the ILD to determine data combinability for slow *Loris* phylogeny. *Systematic Biology*, **50**, 408–424.

Accepted 27 September 2001